Seeking innovation

Deja, Marek; Isto Huvila, Huvila; Widén, Gunilla; Ahmad, Farhan

Published in:
Heliyon

DOI:
10.1016/j.heliyon.2023.e14689

Published: 01/04/2023

Document Version
Final published version

Document License
CC BY

Link to publication

Please cite the original version:
Deja, M., Isto Huvila, H., Widén, G., & Ahmad, F. (2023). Seeking innovation: The research protocol for SMEs' networking. *Heliyon*, *9*(4), Article e14689. https://doi.org/10.1016/j.heliyon.2023.e14689

Heliyon 9 (2023) e14689

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Seeking innovation: The research protocol for SMEs' networking

Marek Deja [a,*], Isto Huvila [b], Gunilla Widén [c], Farhan Ahmad [d]

[a] Jagiellonian University, ul. prof. Stanisława Łojasiewicza 4, 30-348, Kraków, Poland
[b] Uppsala University, Engelska parken, Thunbergsvägen 3H, 752 38, Uppsala, Sweden
[c] Åbo Akademi University, Tuomiokirkontori 3, 20500, Turku, Finland
[d] University of Turku, Rehtorinpellonkatu 3, 20500, Turku, Finland

ARTICLE INFO

Keywords:
Innovation
Knowledge sharing
Natural language processing
Networking
Open innovation
Small- to medium-sized enterprises

ABSTRACT

The paper aims to state the research protocol for the innovation-seeking behavior of Small- to Medium-sized Enterprises (SMEs), related to the classification of knowledge needs expressed in the networking databases. The dataset of 9301 networking offers as the outcome of proactive attitudes represents the content of the Enterprise Europe Network (EEN) database. The data set has been semi-automatically obtained using the rvest R package, and then analyzed using static word embedding neural network architecture: Continuous Bag-of-Words (CBoW), predictive model Skip-Gram, and Global Vectors for Word Representation (GloVe) considered the state-of-the-art models, to create topic-specific lexicons. The proportion of offers labeled as Exploitative innovation to Explorative innovation is balanced with a 51%–49% proportion. The prediction rates show good performance with an AUC score of 0.887, and the prediction rates for exploratory innovation 0.878 and explorative innovation 0.857. The performance of predictions with the frequency-inverse document frequency (TF-IDF) technique shows that the research protocol is sufficient to categorize the innovation-seeking behavior of SMEs using static word embedding based on the description of knowledge needs and text classification, but it is not perfect due to the general entropy related to the outcome of networking. In the context of networking, SMEs place a greater emphasis on explorative innovation in their innovation-seeking behavior. They prioritize smart technologies and global business cooperation, whereas current information technologies and software are more of interest to SMEs that adopt an exploitative innovation approach.

1. Introduction

While observing the growing interest in the issues of digital transformation of enterprises nowadays, it is worth noting that the internet has been, for many years, the principal digital environment in which many small- to medium-sized enterprises (SMEs) have sought information, as it offers 'less expensive access to markets and information on the competition, the economy, and its environment' [1]. However, what distinguishes current business digital communication on the internet is how the present discourse on the digitization of business processes particularly emphasizes the importance of inter-organizational relationships based on knowledge (not necessarily human, also superficially created by machines) that are increasingly necessary to achieve business success [2].

In companies, when individuals organize knowledge management, they may identify gaps in their knowledge and initiate a

* Corresponding author.
E-mail addresses: marek.deja@uj.edu.pl (M. Deja), Isto.Huvila@abm.uu.se (Isto Huvila), gunilla.widen@abo.fi (G. Widén), farhan.ahmad@utu.fi (F. Ahmad).

https://doi.org/10.1016/j.heliyon.2023.e14689
Received 19 December 2022; Received in revised form 14 March 2023; Accepted 15 March 2023
Available online 21 March 2023
2405-8440/

response, which can take the form of human behaviors and innovation-centered processes [3]. These processes involve seeking external ideas to drive positive and successful innovation and fill the knowledge gaps within the company [4]. This outward-oriented process of seeking new knowledge and ideas is referred to as innovation-seeking behavior or simply innovation-seeking in this paper. It is conceptualized as a form of information-seeking behavior, drawing also from the concept of open innovation [5]. Open innovation refers to the strategy of seeking information through collaboration and partnerships with other organizations, institutions, and experts, which companies adopt when they are aware of their limitations [6].

Entrepreneurship involves recognizing and developing innovative opportunities, which requires openness to new ideas and technologies [7]. To remain competitive, entrepreneurs must constantly seek out knowledge and solutions to innovate as a company [8]. The behavior of seeking innovation involves exploring new ideas and taking new collaborative approaches with experts, institutions, and organizations [9–11]. In other words, innovation-seeking stems from being aware of the value of innovation in local and global networks [12]. By keeping an open mind and actively seeking out new information and resources, entrepreneurs can capitalize on innovative opportunities to create successful new products, services, or business models, ultimately determining their success in the fast-paced world of entrepreneurship [13,14].

However, it can be challenging to identify relevant information in the vast amounts of data available, especially in the narrow contexts of different types of innovations in various fields. Misinterpretation and misclassification of an information could lead to the omission of potentially valuable entrepreneurship opportunity [15]. That is why supporting the process of seeking innovation through Natural Language Processing (NLP) is important. Natural Language Processing (NLP) can assist entrepreneurs in analyzing and classifying vast amounts of information to identify relevant insights and opportunities. NLP techniques, such as sentiment analysis and text classification, allow for extracting subjective information from language. Additionally, topic modeling enables entrepreneurs to discover abstract topics in a collection of documents or databases [16]. Using Natural Language Processing (NLP) in information management means that conventional machine learning algorithms, unique NLP methods, and deep learning algorithms are employed to extract big data information and understand documents that are stored in extensive collections. Ultimately, this techniques can significantly enhance the efficiency and accuracy of work [17]. Overall, NLP can be a powerful tool for entrepreneurs to make sense of complex and diverse data sources and facilitate innovation-seeking.

This paper aims to provide an outline of the idea of innovation-seeking behavior and a method to study it at an inter-company level through SMEs' knowledge needs, expressed in networking databases such as the Enterprise Europe Network (EEN). Based on these concepts, it is possible to formulate a research protocol for exploring the content of the innovation-seeking proposals available in the EEN database by using deep learning techniques to recognize and classify the types of innovation sought. In doing so, the paper will contribute to the area of work-related information-seeking behavior by defining innovation as a specific type of knowledge and proposing a natural language processing (NLP) method to study it in an empirical setting. The paper focuses on the following research questions:

1. RQ 1: How to effectively diagnose and categorize massively expressed knowledge needs among companies seeking innovation in networking?
2. RQ 2: What categories of innovation are SMEs likely to seek using a networking database?

In the scope of these two research questions, we make the following contributions:

- We propose a theoretical framework of innovation-seeking behavior related to knowledge needs in the networking of SMEs that was then operationalized using the word embedding technique for the text classification of networking offers.
- Based on the idea that the protocol should be suited for the general use of the biggest networking database, the Enterprise European Network, we compile a code in R language to use simple, most accessible methods for data gathering and word embedding, using static state-of-the-art (SOTA) models (Word2Vec and GloVe) and the TF-IDF technique for baseline comparison.
- We establish a protocol for the categorization of innovation-seeking behavior in networking offers and performance measurement of text classification with the k-nearest-neighbors technique and TF-IDF normalization.
- The research protocol and code published on the osf.io platform are suitable for practitioners focused on content analysis and business scanning for new networking opportunities with a specific approach to innovation.

2. Literature review

A two-part literature review was incorporated to answer fundamental questions regarding our research goal: "What types of innovation can be incorporated into NLP text classification to support innovation-seeking?", "What is the core idea behind innovation-seeking that establishes cooperation and partnership in networking?", and "How networking can be captured through NLP techniques to forecast different innovation-seeking behavior?". The first part was a general literature review, which, according to Grant and Booth (2009), should be performed to capture the most recent or current literature and provide a general topic overview related to our goal [18]. This was necessary because a systematic review did not return any significant results in a search query. The second stage focused on supplementing this overview with a rapid review [18], which was systemized using the CIMO (Context, Intervention, Mechanism, Outcomes) search framework [19]. NLP methodology was omitted in the query as the literature in this regard was captured in the first stage and was very limited.

The query was introduced to the advanced search in Scopus database and structured as follows: ("SME" OR "SMEs") AND (("seeking " AND " innovation") OR "knowledge" OR "R&D") AND ("networking" OR "Enterprise Europe Network") AND ("cooperation" OR "opportunity recognition" OR "partnership"). After narrowing the search strategy chronologically to publications from 2011 and by English language only, the search query returned 32 publications. After screening and rejecting irrelevant publications to our topic, only 24 publications were left after two steps of literature review [10,14,20–43].

2.1. Types of innovation and SME needs

If we define knowledge as a learning outcome that includes acquiring useful information and work-related experience, as well as improving effectiveness and achieving common understanding of success, goals, and efficacy [20], innovation can be defined as "knowledge of the future actions that have been not yet observed" [21]. Knowledge is a combination of experiences, values, and information, as well as expert insights into one's own developments and deficiencies that motivate the search for new experiences and information [22]. Once a knowledge gap is identified, the way the unknown innovation will be understood in each organization will depend on its current state of knowledge, rather than the entropy of future actions. Therefore, a person with knowledge needs can define their knowledge gap through a seeking act that forecasts the need for innovation in the context of the knowledge-seeking process - by describing the new knowledge need, rather than innovation itself. This is because innovation is the result of developing a system of meanings, and the system is not yet ready to classify it because the relevant word clusters do not exist (for the entity, but they can exist in general). Therefore, the most similar applicable category is typically used to classify gaps in current knowledge [21].

In the context of work-related information behavior, the most rational typology that can be embedded in knowledge-based activity is the one operationalized in the study by Ahmad, Huvila, and Widén (2020). They studied the relationship between the information literacy of CEOs and exploratory and exploitative innovation in SMEs. Exploratory innovation departs from existing practices, systems, and markets. It could be the implementation of new solutions and technology that could lead to the entrance into new markets. This is a process where proactive behavior in innovation-seeking is particularly valuable. Exploitative innovation is more incremental and focused on adding value to the current state of companies' development. It brings efficiency into existing practices, enriches current organizational knowledge, and strengthens current processes and structures within a company [23].

Caterina Muzzi and Sergio Albertini (2015) have studied what kinds of competencies are required to support the ability to exploit new knowledge. They show that "prior experience is not inherently necessary to invest in new knowledge domains" [24], so observation of people's behavior as such could be just partially effective while diagnosing innovation-seeking behavior. However, the approach to using new knowledge can be considered as a construct to signify and express the knowledge gaps and knowledge needs related to exploitative and exploratory innovation. In the context of organizational behavior, Esteve Almirall and Ramon Casadesus-Masanell [25] described issues of discovery and divergence as opposite attributes of an open approach to innovation that might bring balance between exploratory and exploitative innovation in the innovation-seeking behavior of SMEs (Fig. 1).

Divergence occurs when the seeker loses the freedom to establish the technological or systemic capabilities of the internal environment. This loss of control can be costly as it accumulates while operating under knowledge gaps that could not have been avoided through the exploitation of one's resources. Divergence brings balance to exploratory and exploitative innovation goals since, in general, the knowledge needs of a partner company and the knowledge gap of a seeker will not be perfectly synchronized (Fig. 1). In such cases, balance is achieved through the discovery of knowledge that needs to be presented in an external environment network. Since suppliers or complementors are generally engaged in product development paths that are different from those presented by the seeker, partners are likely to innovate in ways that the network and seekers have not yet chosen by following their own knowledge needs. Therefore, exploring knowledge needs in a network may enable the seeker to discover new combinations of knowledge in the

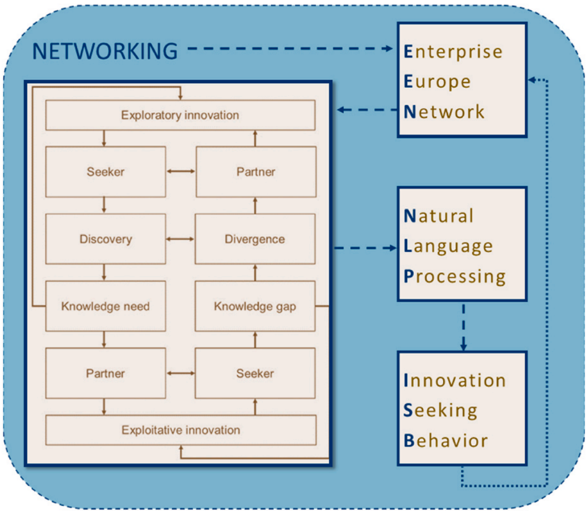

Fig. 1. Identifying the types of innovation-seeking behaviors in networking.

network that would otherwise be difficult to identify through the exploitation of internal gaps [25].

To find both exploratory and exploitative innovation, balance is required between the benefits of new knowledge combinations and the costs of internal coordination related to divergent objectives and cognitive schemas of the two parties in network communication [10,26,27]. Innovation-seeking behavior will be driven by the need to combine "old" knowledge with new knowledge outside the organization where the seeker operates [28]. That is why innovation-seeking behavior will lead to a mix of discovery and divergence in networking, towards both exploitative and explorative innovation. Networks provide SMEs with access to social resources that encourage both exploration and exploitation activities [27]. In our case, the EEN database will also provide a description of knowledge needs related to these resources.

2.2. Seeking innovation

Due to the aforementioned attributes of innovation-seeking, the search for innovation faces natural barriers based on the cognitive distance between the parties involved in the innovation process (i.e., seeker and partner) [14]. Networking can bridge this distance by requiring parties interested in innovation to express and describe what they can and cannot do, thus identifying the gap they want to fill based on their current knowledge. However, divergent needs and patterns of innovation can create complexity in network interactions [26]. Najda-Janoszka and Kopera (2014) discussed barriers in processes similar to innovation-seeking and concluded that most SMEs do not suffer from a lack of new ideas, but rather from the complexity of translating those ideas into comprehensive innovations [29]. Therefore, prior related knowledge is essential for catalyzing both exploratory and exploitative innovation, but only if combined with external knowledge that is at a larger cognitive distance to potential partner [28]. In innovation-seeking behavior, this issue arises in the transition from a knowledge gap to a knowledge need that must be expressed during networking, which precedes cooperation and the creation of innovations of a given type.

In a study analyzing search strategies in innovation-seeking of SMEs, Kim et al. (2020) identified a possible issue that SMEs encounter when transitioning from being aware of the knowledge gap to taking action in the search for new knowledge needs. They defined two dimensions of dealing with this problem as balanced and unbalanced search strategies related to internal and external knowledge sources. The study found that pursuing a balanced knowledge search strategy can lead to the improvement of innovation performance, while a one-sided, unbalanced strategy can weaken a firm's competitive position. The study also found a positive association between organizational ambidexterity and performance, as well as an external-oriented, balanced search strategy [30]. Overall, the issue that is also addressed by our research protocol is that external orientations in networking with a balanced search strategy can minimize institutional distances between SMEs from different regions and business sectors. Supporting this process is now crucial for effective networking scanning [31].

Almirall and Casadesus-Masanell (2010) have pointed out a methodological limitation of their research on networking resulting from the lack of inclusion of all community members and from the fact that, over time, the network focused on innovation shows a very dynamic composition [25]. It is necessary to observe the outcome of network activity, e.g., in a selected networking database, where organizations can actively seek innovation by describing the state and gaps in their knowledge. Therefore, the language patterns describing their needs for new knowledge are the basic two vectors that will indicate their focus on explorative and exploitative innovation.

In order to support innovation-seeking in networking databases using text classification protocols, we should align our word vectors regarding networking with concepts that also aim to minimize the different tensions that companies encounter during the search for innovation opportunities, such as internal and external push and pull motives related to divergent needs of SMEs, which Urbaniec and Żur (2021) described similarly in the context of corporate accelerators. Exploratory innovation-seeking is connected to external push and pull factors, such as new technology development or global collaboration, and exploring new markets or demand transitions. Exploitative innovation-seeking is connected to internal push and pull factors, which covers a range of issues related to efficiency and excellence, limited innovation and internal R&D, providing new information, as well as knowledge acquisition in the form of hiring new talents, organizational learning, new practices, systems, and frameworks [32].

2.3. Networking

Networking refers to a business or R&D network that involves inter-organizational collaborations aimed at innovation and value creation. Its objective is to connect organizations with different assets and competencies, making them more informed, knowledgeable, and responsive to new opportunities [33,34]. According to Jansen, Van Den Bosch, and Volberda's [35] concept of innovation, the key to defining innovation is the cognitive distance from current work practices (competencies), products, customers, and markets, which are mostly assets. The goal of networking is to reduce this distance, both in terms of exploitative and exploratory innovation [24]. Exploitative innovation is strongly associated with current knowledge and focuses on filling knowledge gaps. Therefore, any information resource supporting networking should contain a lexicon of words that accurately describes the exploitative-oriented cluster of knowledge gaps found in any networking database. On the other hand, networking databases are designed to create new value in companies by allowing them to explore new technological possibilities and search for innovations through knowledge needs that can change their environment.

Cooperation and partnerships form the basis for the development of enterprise networking, which is a key element in innovation-seeking. The development of innovation in SMEs is difficult due to the lack of financial resources, the limited opportunities to hire experts, and the general small innovation portfolio in this sector of the economy [24]. Developing the innovation capacity of small businesses requires the participation of other organizations in the knowledge development process. It may also be necessary to include outsiders in R&D initiatives and to involve other organizations in the implementation of innovations [33]. The significance of our study lies in supporting the process of seeking innovation, which can benefit from the exploitation and categorization of networking channels, such as the Enterprise Europe Network (EEN). The networking database enables SMEs to seek partnerships by posting open invitations for collaboration, thus supports creation of a language corpus that can be analyzed by NLP algorithms to generate word vectors that describe the two types of innovation-seeking behavior defined in our research protocol.

3. Related work

Advanced machine learning and NLP methods are rarely used to identify the behaviors and knowledge needs of companies seeking innovation. No studies so far have focused on the use of NLP to identify the innovation-seeking that results from networking initiatives.

3.1. Word embedding and language of innovation

Word vectorization for text classification is a highly desirable area of research, where the reproducibility and applicability of the research framework are crucial, especially in topic- specific sentiment lexicon embeddings related to our research. According to Song et al. [36], unsupervised word embeddings are generally useful, but they may be inadequate for task-specific sentiment lexicon embeddings. The main drawback is the lack of supervision between a word and its associated contexts [37].

The unsupervised nature of word embeddings is a significant disadvantage, but it actually benefits our research protocol focused on networking and innovation. This is because the knowledge needs in this context don't have a simple polarity between exploratory and exploitative innovation, which means we don't need to apply an objective function to optimize word vector sentiment labels [38,39]. While external input knowledge may improve category prediction, we aim to keep the research protocol as simple as possible. Enhancements can be left for future projects and implications of this study within the context of networking and R&D innovation processes.

3.2. Static word embedding models in the context of innovation

The studies most related to our goal have mostly focused on product innovation and user reviews [40], technological opportunities and user needs [41], and the language of innovation in technological patent descriptions [21], as well as patent semantic analysis [42,43]. Although Natural Language Processing (NLP) provides direct evidence of innovation-seeking, it has its limitations due to the lack of proximity to the people creating these offers. Analytical techniques like classification, topic modeling, and text classification are commonly used in research to identify themes, similarities, and differences between documents and the topic orientation of texts [17]. However, this still leaves a gap for exploration regarding the semantics of innovation in companies.

4. Methodology

In this paper, we do not want to develop a new deep learning-based model or model augmentation, but to present a protocol for the use of well-proven SOTA models for a new research goal, in a previously not considered context of networking database content. The proposed framework includes three stages.

- Firstly, we want to perform a semi-automated data gathering using a web scraping technique in the publicly accessible database – Enterprise European Network maintained by the European Commission.
- Secondly, we have used the static word embedding techniques to transform the EEN database vocabulary into dense vectors of real numbers in the language corpus from which we want to develop topic-specific lexicons to understand networking in the context of two types of innovation-seeking behavior.
- Thirdly, by conducting lexicon wordcount on sample offers from the EEN database it was possible to categorize and label their content, respectively, to each of the two innovation-seeking behaviors. In this step, we want to evaluate our categorization with the machine learning k-nearest neighbors classification technique and compare the prediction performance of each lexicon using the TF-IDF technique for baseline comparison.

4.1. Research protocol

The proposed research protocol aims to utilize NLP algorithms to analyze and classify the content of the networking database based on the binary innovation division of exploratory and exploitative innovation. Machine learning will be used to evaluate the classification performance, based on how accurately the predictions of innovation-seeking behavior can be made [44].

Our research procedure is designed to provide us with the material needed to answer the following detailed questions related to RQ 2:

- RQ 2a: What is the proportion of word counts in the EEN offers that represent the two orientations of innovation-seeking behavior, i.e., exploratory and exploitative innovation?
- RQ 2b: Which type of innovation is emphasized more in the EEN networking database?

- RQ 2c: What is the precision of the prediction of each type of innovation, and which type can be predicted more accurately based on the content of the EEN database?
- RQ 2d: What are the differences in the subject matter of the offers classified as exploratory vs. exploitative innovation?

4.1.1. Enterprise European Network

To begin, we selected content related to innovation-seeking in networking, using the EEN database. This initiative by the European Commission aims to help companies innovate and expand their reach globally. The offers published on the platform are intended to facilitate knowledge transfer between organizations, allowing them to manufacture products more effectively, access new markets more easily, find the technology they need to drive innovation faster, and collaborate more openly in research and development projects. From the perspective of our research agenda, the EEN database is a valuable source of information, with about 3.5–5.5 thousand equally structured networking offers being published by companies on an ongoing basis. The content of these offers varies over time, as no archive is available, but they all pertain to buying, selling, teaching/learning, implementing, and collaborating (in R&D projects), all in the context of exploiting new knowledge within the network. The database contains a wealth of information that describes the knowledge needs of many SMEs, as well as the outcomes of their innovation-seeking behavior.

4.1.2. Data collection

We use the *rvest* package in R language to take data from the EEN database [45] and *the word2vec* and *text2vec* packages to perform word embeddings [46,47]. Our code obtained text data from 9301 networking offers through CSS nodes in the EEN database (https://een.ec.europa.eu/partners). We aim to collect as many offers as possible to gain the best representation of SMEs' needs, so we do not apply any filters at this stage. However, during the data analysis phase, we establish a selection criteria to include only those offers with high saturation of words from automatically created lexicons.

4.1.3. Networking dataset

Two datasets were used during the two stages of the study. In the first quarter of 2021, we gathered 5635 offers from the Enterprise European Network database and created a topic-specific corpus with 95,110 sentences and 11,270 unique words. In the second stage, a year later, we collected additional 3666 offers and added them to create a corpus of 9301 networking offers, 157,133 sentences, and 18,602 unique words related to networking. We collected additional offers and conducted the second stage of the study because the first corpus was too small to achieve fully satisfactory results. Based on that corpus, we used static word embedding models described below to create topic-specific lexicons that are part of our dataset used in text classification.

The entire procedure for obtaining data, developing lexicons, and performing text classification in the R language was included in the project on the OSF platform: https://osf.io/qa5rm/.

4.1.4. Static embedding models

Choosing between static and dynamic embedding models is a key issue in addressing content related to networking and innovation-seeking [48]. Networking, as a form of behavior among SMEs, focuses on finding partners, and has a consistent goal despite the many different industries that initiate business contacts in such networks. The approach to innovation and the search for knowledge is not polarized, but rather a mixture of behaviors, and the dominance of one behavior would be the basis for categorization. Therefore, partial overlapping of lexicons is not a problem. As a result, static word embedding models are sufficient for recognizing the type of behavior related to the innovative approach of SMEs, and categorization may present the general nature of the knowledge need depending on the behavior.

4.1.5. Text classification

Topic-specific text classification is a well-known and commonly used method of content analysis. It includes various methods that allow researchers to determine the polarity of a sequence of words in a text based on its lexical corpus. For example, positive versus negative orientations in sentiment analysis, or topic-specific dictionaries such as populism versus liberalism orientations [49].

In our study, we use exploratory versus exploitative innovation as labels to categorize each networking offer and create an output of our lexicon-based text classification. To do this, we employ three common unsupervised static word embedding algorithms to build simple topic-specific binomial lexicons containing words associated with the two types of innovation in our framework. This method is applicable only to the EEN database.

Based on the literature review, we have selected three state-of-the-art models: CBOW and Skip Gram from the Word2Vec toolbox [46,50] and GloVe [51]. The Skip-Gram and Continuous Bag-of-Words (CBoW) models are neural networks trained to predict the distributional embedding of dense word vectors. CBoW is a type of neural network that predicts the probability distribution of words in a context given by a surrounding word window. It is used specifically in the field of word embedding to predict a target word based on the context of the surrounding words. It takes a sequence of words as input and produces a single word as output. Skip-gram, on the other hand, is another neural network architecture used for generating word embeddings. Unlike the CBoW model, Skip-gram is a predictive model that takes a target word and tries to predict the surrounding context words. It can be trained on large amounts of text data to maximize the probability of observing the context words within a fixed-size window around the target word [46,50].

The GloVe model can be used to calculate the co-occurrence matrix for key phrases. GloVe uses co-occurrence statistics of words in a large corpus of text to create word embeddings that capture semantic relationships between words. Global Vectors (GloVe) analysis is related to the Word2Vec method of Mikolov et al. [50] and the factorization of word co-occurrence matrices [51]. In Word2Vec, the vectors are derived from a classification task that relies on contextual word co-occurrences. The GloVe model, on the other hand, is trained on general co-occurrence statistics.

The deep learning networks using the above-mentioned models were used to obtain associations of vocabulary in the EEN lexical corpus with their typical contexts and symbolic aspects of language related to exploratory and exploitative innovations (two language clusters). The final step was to use an R script to sort the 200, 400, and 600 words that are most related to terms for exploratory innovation and exploitative innovation. Based on the scoping review of the innovation context, we separated exploratory innovation by creating word vectors for the terms: innovation, knowledge, explore, develop, create, design, technology, new market, external, -exploit, and exploitative innovation with terms: innovation, information, exploit, change, acquire, know-how, process, system, internal, and -explore. We used cosine similarity as the most widely used metric to calculate semantic similarities between word embeddings [48]. The results of these transformations are 18 lexicons separated by three algorithms, three levels of word numbers, and two types of innovation-seeking behavior (Fig. 2).

Through text classification, we were able to identify words related to two types of innovation, and using the z-score method, we assigned scores to each offer based on their innovation type [52]. The z-scores of each document statistic indicate which type of innovation-seeking behavior, either exploratory or exploitative, is closer to the mean value of the word count in the sample. To assess the accuracy of this categorization, we used nonparametric statistics and the k-nearest classification techniques [53].

In order to evaluate and compare the results of text classification across each lexicon, we conducted a baseline comparison using the term frequency-inverse document frequency (TF-IDF) technique. TF-IDF has previously been used effectively as a feature extraction technique for generating a set of feature vectors for baseline comparison [54]. This technique reflects the importance of a word to a document in a collection or corpus of documents. TF-IDF increases the weight of terms that are specific to a single document and decreases the weight of terms used in most documents, making it widely used in sentiment analysis [55]. It is suitable for creating a

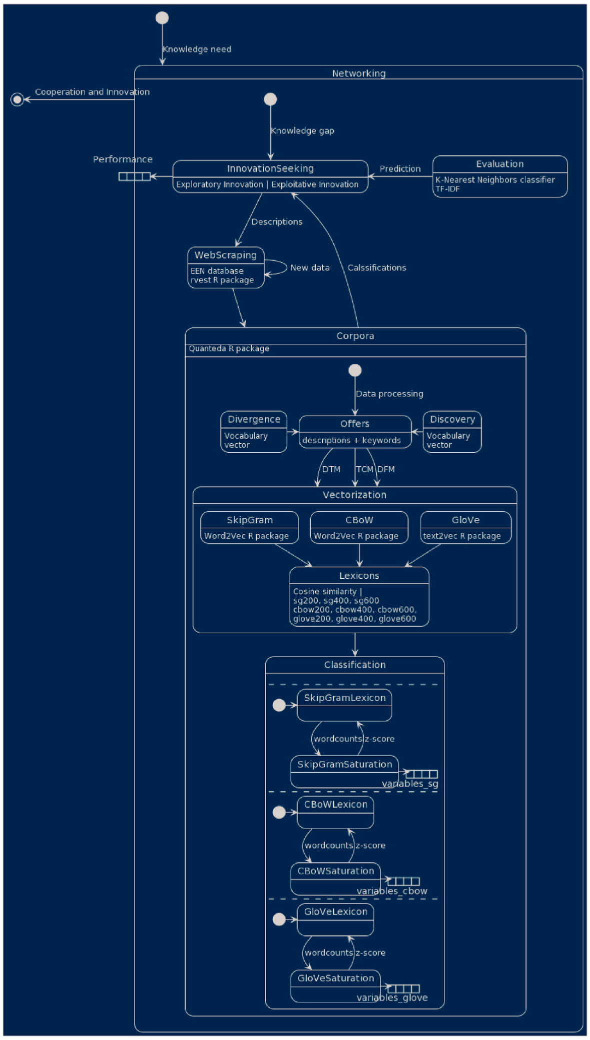

Fig. 2. Research protocol.

benchmark for feature representation methods that can compare the prediction performance of our data model for each text classification output, which results in a labeled dataset.

4.2. Analysis

To begin the analysis, we first identified outliers - observations (i.e., offers) that were insufficiently saturated with the lexicon under study. This was necessary because we recognized that, in some cases, networking may not necessarily focus on innovation but rather on the need to purchase new products. Therefore, if the text classification showed a small proportion of innovation lexicon saturation, then these offers would not be relevant for exploration and exploitation innovation classifications as we and the algorithm understand them. The resulting test sample consisted of 3770 unique offers published in the EEN database (n = 3770).

4.2.1. Nonparametric statistics analysis

The second step involved comparing the median word count values in each lexicon for the selected sample, as shown in Table 1. Overall, the word count for exploitative innovation was found to be approximately 20–30% higher than that for explorative innovation. Specifically, for the 200-word lexicons based on the Skip Gram, Bag of Words, and GloVe algorithms, the number of words returned for explorative innovation were 48, 80, and 66, respectively, while the numbers for exploitative innovation were 33, 46, and 46, respectively. For the 400-word lexicons, the Skip Gram, Bag of Words, and GloVe algorithms returned 105, 121, and 131 words for explorative innovation, and 78, 82, and 99 words for exploitative innovation, respectively. Lastly, for the 600-word lexicons, the Skip Gram, Bag of Words, and GloVe algorithms returned 126, 159, and 158 words for explorative innovation, and 69, 119, and 149 words for exploitative innovation, respectively.

The second step of our analysis aimed to evaluate the effectiveness of the z-score-based classification in text labeling for each lexicon used. We used the paired sample nonparametric Mann-Whitney *U* test to compare the proportions of each innovation type identified by the classification (Table 2). Nonparametric tests are commonly used to evaluate continuous variables from NLP text classification because they often have nonparametric distributions, issues with equality of variances, and potential outliers [56]. Our research question 2a is linked to the main alternative hypothesis of this test, which is that the two samples are not equal [57,58]. If the p-value of the test is very low, we can suspect that the difference in metric is not random [59].

We used the Mann-Whitney *U* Test to initially compare the performance of Skip Gram, CBoW, and GloVe algorithms in lexicon-based text classification and determine if there were significant differences in word counts between the categories of innovation. Our results showed that explorative innovation had a statistically significantly higher median word count than exploitative innovation. Furthermore, the independent sample *U* Test confirmed that exploitative innovation was always represented by a statistically significantly lower median word count than explorative innovation ($p < 0.001$).

Our analysis aims to demonstrate the validity and accuracy of our predictions regarding different types of innovation in EEN offers, using a trained algorithm in text classification (Table 3). We evaluated the separation accuracy of each lexicon using the procedure developed by Alexander Ly and Koen Derks [60], and found that the accuracy is consistently high (AUC >0.94). However, this evaluation only measures the accuracy of our automated categorization, based on output labels from our dataset. We don't have information about the actual networking results, but since our classification is well-balanced in terms of innovation types, AUC is an appropriate evaluation metric.

To make predictions, we used the TF-IDF normalization technique on the entire dataset, with a 20% test set indicator ($n_1 = 1127$; $n_2 = 1860$) and 80% training set ($n_1 = 4508$; $n_2 = 7441$) (Table 3). In the first stage, CBoW and GloVe algorithm-based lexicons performed well (AUC > 0.850). Skip Gram lexicons showed the biggest difference in accuracy of classification for distinguishing between Exploratory and Exploitative innovation, while there was no significant difference in CBoW lexicons (CBoW 400).

Table 1
Descriptive statistics for text classification.

	N	Median	Mean	SD	SE
sg200.explorative	3770	48	52.238	22.791	0.371
sg200.exploitative	3770	33	37.825	21.597	0.352
sg400.explorative	3770	105	114.849	47.419	0.772
sg400.exploitative	3770	78	86.506	38.502	0.627
sg600.explorative	3770	126	134.778	49.258	0.802
sg600.exploitative	3770	69	78.055	39.236	0.639
cbow200.explorative	3770	80	84.691	29.524	0.481
cbow200.exploitative	3770	46	49.160	18.091	0.295
cbow400.explorative	3770	121	127.030	41.277	0.672
cbow400.exploitative	3770	82	87.326	31.137	0.507
cbow600.explorative	3770	158.5	167.430	56.590	0.922
cbow600.exploitative	3770	119	127.125	46.839	0.763
glove200.explorative	3770	66	69.856	26.152	0.426
glove200.exploitative	3770	46	49.160	18.091	0.295
glove400.explorative	3770	131	138.789	45.487	0.741
glove400.exploitative	3770	99	105.719	39.019	0.635
glove600.explorative	3770	158	166.920	55.117	0.898
glove600.exploitative	3770	149	159.106	54.715	0.891

Table 2
Independent samples Mann-Whitney *U* test in Text classification.

				95% CI for Rank-Biserial Correlation						
	W	P	Rank-Biserial Correlation	Lower	Upper	Group	N	Mean	SD	SE
SkipGram200	842512.000	<.001	−0.524	−0.551	−0.497	Exploitative	1779	36.559	14.970	0.355
						Explorative	1991	49.673	16.166	0.362
SkipGram400	846212.000	<.001	−0.523	−0.549	−0.496	Exploitative	1963	82.423	33.464	0.755
						Explorative	1807	118.806	45.792	1.077
SkipGram600	339296.000	<.001	−0.809	−0.821	−0.796	Exploitative	1811	74.369	33.596	0.789
						Explorative	1959	136.810	42.900	0.969
CBoW200	**168829.500**	**< .001**	**−0.905**	**−0.911**	**−0.898**	**Exploitative**	**2006**	**45.733**	**14.796**	**0.330**
						Explorative	**1764**	**86.200**	**22.320**	**0.531**
CBoW400	463856.000	<.001	−0.739	−0.755	−0.722	Exploitative	1910	83.548	27.667	0.633
						Explorative	1860	129.165	35.028	0.812
CBoW600	854740.000	<.001	−0.519	−0.545	−0.491	Exploitative	1879	125.284	42.432	0.979
						Explorative	1891	166.818	49.733	1.144
GloVe200	345077.000	<.001	−0.804	−0.817	−0.791	Exploitative	2059	45.133	14.004	0.309
						Explorative	1711	71.243	17.802	0.430
GloVe400	764886.500	<.001	−0.569	−0.593	−0.543	Exploitative	1957	102.774	34.506	0.780
						Explorative	1813	140.065	40.260	0.946
GloVe600	1.635e+6	<.001	−0.079	−0.116	−0.042	Exploitative	1845	158.748	50.975	1.187
						Explorative	1925	165.209	49.667	1.132

Note. For the Mann-Whitney test, the effect size is given by the rank biserial correlation.
Note. Mann-Whitney *U* test.
Very good separation was observed in the case of CBoW 200 words lexicons with 81% of the observations confirming a statistically significant difference in word count between the two types of innovation to the benefit of Explorative innovation ($n = 1764$, Mdn = 81, $\bar{x} = 86.2$, SE = 0.531). Also good separation was observed in case of Skip Gram 600 words lexicon ($n = 1959$, Mdn = 127, $\bar{x} = 136.81$, SE = 0.969), CBoW 400 words lexicon ($n = 1860$, Mdn = 122, $\bar{x} = 129.165$, SE = 0.812), and GloVe 200 words lexicon ($n = 1711$, Mdn = 68, $\bar{x} = 71.243$, SE = 0.43).

Overall, our experiment with a sample of 9301 offers showed significant improvement in all measures, with the best performance shown by CBoW-based and GloVe-based lexicons. While the GloVe400-based categorization had a slightly better AUC score (0.888 vs. 0.851), the CBoW400 categorization had the highest F1 score (0.851 vs. 0.774), which could be important when dealing with imbalanced data in the future. Using CBoW and GloVe algorithm-based lexicons, we achieved precision rates of 89% and 88% for Exploitative innovation, and 86% and 81% for Explorative innovation, respectively.

5. Results

Based on the qualitative analysis of the ROC curves (Fig. 3) and the prediction performance evaluation metric above (Table 3), it is the CBoW-based lexicons that have shown slightly better performance in predicting the type of innovation in the EEN labeled dataset. Regarding *RQ 2*, massively expressed needs for innovation and the different categories of innovation-seeking behavior can be very effectively identified by using machine learning and text classification techniques.

Regarding *RQ 2b*, looking at the proportion of words in binary innovation classification concerning the total word count in each offer (Fig. 4) by using the CBoW 400 lexicon, we can measure an average of 25% share of Explorative innovation to 17% share of Exploitative innovation. The median value of innovation word count that is the most similar to the median score of the 18 classification lexicons used can be observed using 400 words of CBoW-based lexicons (Fig. 4; Explorative = 122|112; Exploitative 79|78.5).

It's important to note the topic differences between EEN offers oriented towards Exploratory and Exploitative innovation (Table 4), considering that the general word count in the case of Exploratory innovation is approximately 20–30% higher than that of Exploitative innovation. Additionally, the overall share of innovation-related words is also around 7% higher for Exploratory innovation. Although both types of innovation share topics related to certain industries and services, which may be due to the specificity of the EEN base and its popularity in some market sectors, clear differences indicate a changing concentration of SMEs towards the use of the latest technologies and a focus on international innovation-seeking.

In relation to *RQ 2d*, our analysis of primary word combinations ranked from 1 to 20 based on word counts in our datasets reveals that the first three positions in rankings are similar for documents classified as related to Exploratory innovation and Exploitative innovation. However, there are significant differences between the two types of innovation beyond the top five ranking positions. These differences are mainly related to the orientation towards seeking international business partners and activities in information technology and computer-related consultancy. Specifically, SMEs seeking exploratory innovation are more focused on international business partnerships, while SMEs seeking exploitative innovation are more focused on solutions and consulting related to information technology (Table 4).

Table 3
Prediction performance using K-nearest neighbors and TF-IDF algorithms.

Lexicon	Nearest-neighbors	n (Train)	n (Validation)	n (Test)	Validation Accuracy	F1 Test Accuracy	AUC	Precision rate based on the confusion matrix	
								Exploitative	Explorative
SkipGram200	40	2404	602	758	0.978	0.975	0.998	0.978	0.973
*						0.656	0.706	0.704	0.584
**						**0.756**	**0.862**	**0.621**	**0.760**
SkipGram400	17	1611	403	504	0.945	0.938	0.983	0.936	0.942
*						0.538	0.531	0.588	0.412
**						**0.754**	**0.867**	**0.766**	**0.771**
SkipGram600	17	1611	403	504	0.950	0.956	0.996	0.967	0.946
*						0.694	0.751	0.735	0.632
**						**0.740**	**0.880**	**0.752**	**0.757**
CBoW200	23	1611	403	504	0.965	0.972	0.997	0.974	0.970
*						0.741	0.817	0.762	0.706
**						**0.775**	**0.852**	**0.758**	**0.788**
CBoW400	15	1611	403	504	0.953	0.970	0.996	0.972	0.969
*						0.785	0.851	0.804	0.756
**						**0.851**	**0.887**	**0.878**	**0.857**
CBoW600	24	1611	403	504	0.931	0.954	0.994	0.944	0.964
*						0.753	0.818	0.774	0.719
**						**0.806**	**0.859**	**0.885**	**0.831**
GloVe200	7	1606	402	510	0.965	0.953	0.993	0.962	0.942
*						0.697	0.785	0.749	0.628
**						**0.781**	**0.843**	**0.738**	**0.778**
GloVe400	15	1606	402	510	0.938	0.967	0.996	0.978	0.978
*						0.706	0.785	0.763	0.656
**						**0.774**	**0.888**	**0.884**	**0.812**
GloVe600	29	1606	402	510	0.955	0.959	0.994	0.945	0.973
*						0.688	0.727	0.714	0.642
**						**0.815**	**0.850**	**0.852**	**0.816**

Note. The model is optimized for the accuracy of the validation set.
Note. Weights: rectangular. Distance: Euclidean.
* Performance measures when predicting classification using TF-IDF normalization based on a corpus consisting of 5635 offers ** Performance measures when predicting classification using TF-IDF normalization based on a corpus consisting of 9301 offers.

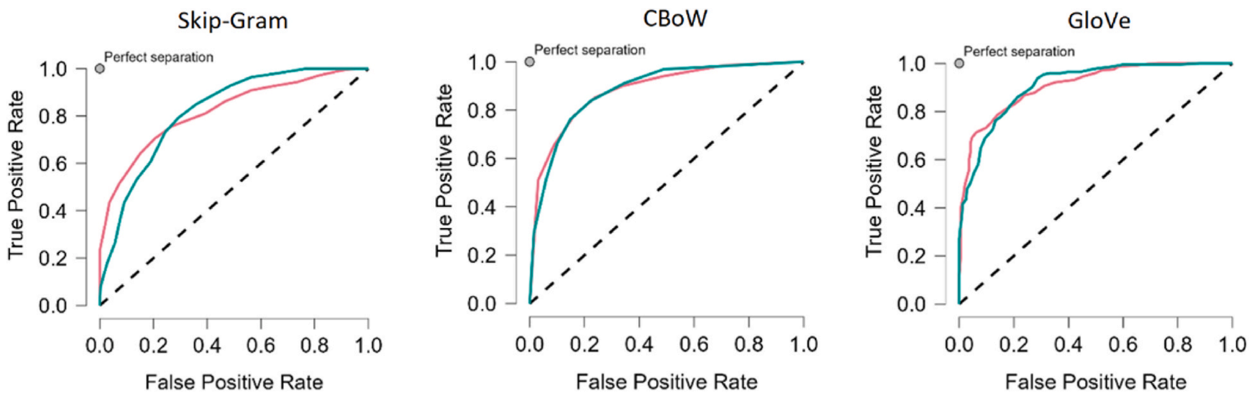

Fig. 3. ROC curves for the innovation text classification results. In turn: Skip Gram, Bag-of-Words, GloVe.

6. Discussion

6.1. NLP practical implications

The results demonstrate that the research protocol is adequate for predicting SMEs' innovation-seeking behavior in networking. However, the prediction performance is highly reliant on the size of the lexical corpus and the number of offers used for its development. Since there is no direct comparison to a text classification study of networking offers, the results should be at least comparable to the prediction performance measures (F1 and AUC) in similar research designs and word embedding models, albeit in different contexts.

Farman Ali et al. [61] conducted sentiment analysis on transportation data using several machine learning algorithms and

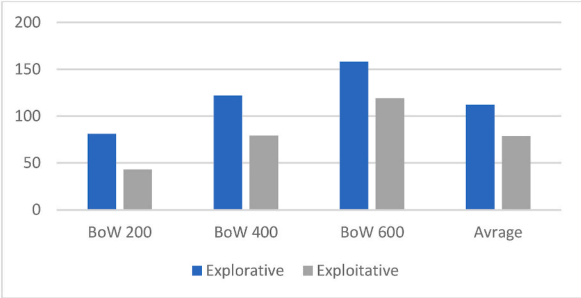

Fig. 4. Average word count of Explorative and Exploitative innovation in EEN offers.

Table 4
The ranking of the main explorative and exploitative innovation contexts occurring in the EEN offers (based on four-word combinations with lemmatization).

Primary word combinations	Exploitative ranked	Exploratory ranked	Rank difference
identify international business partner	26	**7**	19
seek an international business partner	25	**8**	17
activity other information technology	**6**	18	12
activity computer consultancy activity	**9**	18	9
motor vehicle transportation equipment	**12**	20	8
automation robotic control system	18	**11**	7
computer programme activity computer	**8**	15	7
other consumer product manufacture	24	**19**	5
artificial intelligence relate software	11	**6**	5
software artificial intelligence relate	21	**17**	5
machine turn drill mould	6	10	4
other industry specific software	19	22	3
agriculture forestry fish animal	1	3	2
accessory include jewellery manufacture	16	14	2
partner under distribution service	23	21	2
under distribution service agreement	2	1	1
other metal work equipment	3	2	1
other human health activity	10	9	1
medical technology biomedical engineer	12	11	1
other electronic relate equipment	14	13	1

word2vec. Their approach achieved higher precision measures of 90%, comparable to our results. Additionally, when comparing our results with studies on dynamic word embedding, such as Pasupa and Ayutthaya [62], where the authors utilized deep learning techniques for sentiment analysis in the Thai language, they achieved the highest F1 score of 0.817 with the CNN model. Therefore, it can be concluded that our results demonstrate good and comparable prediction performance measures, but they are not perfect.

Furthermore, in the context of dynamic embedding models such as BERT being used in aspect-based sentiment analysis, Ben Liang et al. achieved the highest F1 score of 87% [63]. In the study that is most similar to ours, focusing on mining product innovation by Zhang et al. (2021), the authors used a combination of different dynamic models to perform word embeddings [40]. They achieved very high prediction performance with AUC 0.91 and F1 0.89, which is comparable to our slightly lower results in stage two of the experiment (AUC 0.887, F1 0.851). Taking into account the nature of networking and the diverse needs described in the EEN database, we can conclude that our experiment was successful. While the use of dynamic word embeddings would likely improve performance measures, the improvement might not be significant, and our protocol can be considered a time- and resource-saving solution.

6.2. Theoretical implications

Based on the content analysis performed, it is clear that SMEs oriented towards Explorative innovation are more focused on artificial intelligence-related software and automation or robotic control systems, whereas those oriented towards Exploitative innovation are focused on current solutions for computer programs and equipment related to a specific market, transportation. Regarding RQ 3, we can conclude that SMEs seeking exploitative innovation are more focused on improving processes with the help of new software (new for the seeker) and generally on purchasing solutions and hardware that will strengthen their internal performance.

Innovation is a high-risk venture, and not all companies want to take this risk by only bidding on creative and agile problem solutions. Instead, they choose to enter a viable product, service, or purchase process [64]. However, since the innovation process is strongly supported by the phenomenon of networking, which involves knowledge exchanges between multiple parties, it becomes a creative issue to turn a problem or idea into state-of-the-art products, services, and processes in exploratory-oriented offers that might not be viable in principle as they explore the knowledge gap of the seeker. This issue shifts the boundary of what is feasible now to the entropy related to what will be feasible when the knowledge exchange occurs. Even if such an innovation process does not result in a new high-end product with the intended prior purpose, the innovation-seeking behavior may settle for a new process with completely different valuable knowledge. Therefore, organizations interested in networking could respond effectively to just an innovative effort when their knowledge-absorbing capacity is high, but general creativity or resources are low [65,66].

The effectiveness of our semi-automated categorization of innovation-seeking behavior is based on two fundamental perspectives of the innovation of SMEs in networking in terms of their absorptive capacity in networking: potential or realized absorptive capacity and substitution or evolution of capabilities to innovate [67,68]. Although some major skills and knowledge gaps might appear as triggers of or during the innovation-seeking process, it is also necessary to fulfill these gaps, respectively, to the absorptive capacity of a company, i.e., the general staff's capacity to learn, implement new knowledge, or disseminate new technology. Such limitations will then be a potential cause of having a mixture of different attitudes in the innovation-seeking process. SMEs using networking databases might be at the crossroads between what kind of knowledge they potentially want to find and absorb as a result of innovation-seeking; therefore, each EEN offer includes a detailed description of the status and achievements of each company.

Networking facilitates social interactions among smaller firms and helps them realize their innovation potential. SMEs typically belong to at least one network, and these networks are primarily driven by the exchange of business and technical knowledge [67]. Networking has been found to enhance the potential for exploratory innovation, while exploiting performance improvements and data-driven learning can optimize existing innovation trajectories. Participating in exchange-driven ecosystems can reduce innovation barriers and increase absorptive capacity [69]. Our study indicates that presented protocol effectively measure certain aspects of exploratory and exploitative innovation.

The internal exploitation of knowledge is the final point of the transformation process during innovation for both exploratory and exploitative approaches, but SMEs are mostly considered to be a subject of one of them [67]. Our categorization shows different perspectives where innovation-seeking behavior might be related to differences in absorptive capacity, but it does not induce a perfect duality between exploration and exploitation innovation. As Duy Quoc Nguyen [70] described, exploratory and exploitative innovation 'are complementary in the fit-as-mediation form in which R&D mediates the influence of novel external knowledge'. The shift between exploratory and exploitative focus in innovation-seeking will be conditional on the capacity reconfiguration mechanism of a company, where capacity substitution will be a concern of exploratory innovation-seeking, and capacity evolution might be more of a concern in the case of exploitative innovation-seeking [70].

This study has significant implications for the internationalization of SMEs. The findings indicate that Exploratory innovation is a more prevalent subject of innovation-seeking behavior in the EEN database (Fig. 2), as opposed to an exploitative approach. This suggests a trend in networking preferences where SMEs prioritize intelligent technologies and an international approach to business operations. These perspectives are more relevant to R&D activities than the transformation or evolution of current information technologies and software, which characterizes exploitative innovation. Thus, the original assumptions regarding exploratory innovation hold true for the EEN database as well.

The exploratory innovation-seeking behavior of SMEs is focused on making an impact on the market. It is worth noting the strong emphasis on implementing artificial intelligence in this category, driven by the belief that industry 4.0 technologies, such as machine learning and deep learning, will enhance decision-making and enable the creation of new solutions, ultimately accelerating the exploration and development of useful technologies and services [71]. This approach aligns with the principles of open innovation, where collaboration based on new technologies can not only improve SMEs' efficacy but also lead to changes in the organizational environment, transform operations, and enable progress towards Industry 4.0 [72].

6.3. Limitations of the study

One main limitation of our research protocol is that it was tested based on the implied distance between companies and people seeking innovation. Overall, our study reflects only the perspective of European-centered networking supported by European Commission. While EEN is one of the largest and most accessible networking databases in the world, our observations cannot be generalized to all such initiatives regarding innovation and R&D. Also, it is possible that in cases where exploitative innovation is sought, it may also be more necessary for efficient or sustainable operations, as indicated by Pereira et al. (2019). Furthermore, each offer may contain elements of both explorative and exploitative innovation, and the categorization will depend on which type is more representative. However, the actual exchange of knowledge and transformation within the organization may be opposite to the goal of the seeking process. Therefore, while our research protocol cannot completely eliminate the entropy associated with innovation and future actions of companies, it can still support and optimize the search process through effective categorization and predictions of the type of innovation. The main added value of our experiment is confirming that in the context of networking, one type of innovation is simply more or less emphasized by SMEs during innovation-seeking.

Another unknown is how effective our research protocol would be in databases such as CORDIS or other innovation or R&D-related databases. It's important to consider the implications of repeating this procedure in other databases, on much larger datasets, as well as testing it on the EEN with pre-trained, transformer-based, and dynamic word embedding algorithms, such as LSTM, CNN, or Google BERT.

7. Conclusions

Our study has presented few several advantages in identifying SMEs' innovation-seeking behavior through NLP text classification. One of the main strengths is that our research protocol provides insight into the nature of network communication and enhances the understanding of networking initiatives among SMEs. By analyzing the exchange of business and technical knowledge in these networks, we can predict innovation-seeking behavior and identify types of innovation needs. This information can be valuable to users of networking databases (business analysts, embedded librarians, researchers and entrepreneurs) allowing them to select only the offers that are more likely to result in a specific type of innovation approach. Our research protocol can be openly repeated, enabling others to replicate our findings and further explore innovation-seeking behavior in SME networks. The use of NLP text classification algorithms in this study can be extended to other innovation and R&D-related databases, further enhancing our understanding of innovation-seeking behavior of different entities.

Author contribution statement

Marek Deja, PhD: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Isto Huvila; Gunilla Widén; Farhan Ahmad: Conceived and designed the experiments; Wrote the paper.

Funding statement

Dr Marek Deja was supported by Jagiellonian University as part of the strategic program "The Excellence Initiative – Research University," which also funded the publication of this work under an Open Access license.

Data availability statement

Data associated with this study has been deposited at Open Science Framework, https://osf.io/qa5rm.

References

[1] A. Fernández-Portillo, M. Almodóvar-González, M.C. Sánchez-Escobedo, J.L. Coca-Pérez, The role of innovation in the relationship between digitalisation and economic and financial performance. A company-level research, Eur. Res. Manag. Bus. Econ. 28 (2022), 100190, https://doi.org/10.1016/j.iedeen.2021.100190.
[2] L. Gastaldi, S. Lessanibahri, G. Tedaldi, G. Miragliotta, Companies' adoption of Smart Technologies to achieve structural ambidexterity: an analysis with SEM, Technol. Forecast. Soc. Change 174 (2022), 121187, https://doi.org/10.1016/j.techfore.2021.121187.
[3] S. Yata, F. Hurd, 'Keep Your Friends Close': the role of informality in the management of unequal-sized collaboration for NZ SMEs, J. Asia Pac. Bus. 22 (2021) 84–104, https://doi.org/10.1080/10599231.2021.1866402.
[4] M. Ciesielska, Nokia on the slope: the failure of a hybrid open/closed source model, Int. J. Entrep. Innov. 19 (2018) 218–225, https://doi.org/10.1177/1465750317742843.
[5] U. Lichtenthaler, E. Lichtenthaler, A capability-based framework for open innovation: complementing absorptive capacity, J. Manag. Stud. 46 (2009) 1315–1338, https://doi.org/10.1111/j.1467-6486.2009.00854.x.
[6] A. Cavallo, H. Burgers, A. Ghezzi, V. van de Vrande, The Evolving Nature of Open Innovation Governance: A Study of a Digital Platform Development in Collaboration with a Big Science Centre, Technovation, 2021, 102370, https://doi.org/10.1016/j.technovation.2021.102370.
[7] T. Tuomisalo, Emergence of an entrepreneurial opportunity: a case within a Finnish telecommunication international new venture, J. Int. Entrep. 17 (2019) 334–354, https://doi.org/10.1007/s10843-019-00247-2.
[8] A.J. Guerrero, J. Heijs, E. Huergo, The effect of technological relatedness on firm sales evolution through external knowledge sourcing, J. Technol. Tran. (2022), https://doi.org/10.1007/s10961-022-09931-3.
[9] G.M. Ahimbisibwe, M. Ngoma, A.K. Nabatanzi-Muyimba, L.B. Kabagambe, Entrepreneurial mindset and SME internationalization in Uganda: the mediating role of international networking, Rev. Int. Bus. Strategy (2023), https://doi.org/10.1108/RIBS-11-2021-0149.
[10] T. Paiva, M.P. Ribeiro, P. Coutinho, Capacity-building model to promote innovation and sustainability in the Portuguese agro-industrial sector, Sustain. Switz. 14 (2022), https://doi.org/10.3390/su142315873.
[11] A. Basile, R&D networking Model for support of SME's competitiveness, Int. Bus. Manag. 6 (2012) 417–425, https://doi.org/10.3923/ibm.2012.417.425.
[12] A.B. Memon, K. Meyer, M.N. Tunio, Toward collaborative networking among innovation laboratories: a conceptual framework, Int. J. Innovat. Sci. 14 (2022) 282–301, https://doi.org/10.1108/IJIS-04-2021-0069.
[13] T. Roolaht, Value creation in globalizing SMEs, Value Creat. Int. Bus. 2 (2016), https://doi.org/10.1007/978-3-319-39369-8_2. SME Perspect.
[14] M. Cardoso, I. Ramos, Rvolta, a case for open innovation: how can a SME be innovative in a competitive industrial environment? Small Medium Enterp. Concepts Methodol. Tools Appl. 4 (2013) https://doi.org/10.4018/978-1-4666-3886-0.ch079.
[15] I. Uvarova, A. Vitola, Innovation challenges and opportunities in European rural SMEs, Publ. Pol. Adm. 18 (2019) 152–166, https://doi.org/10.5755/j01.ppaa.18.1.23134.
[16] J. Prüfer, P. Prüfer, Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in The Netherlands, Small Bus. Econ. 55 (2020) 651–672, https://doi.org/10.1007/s11187-019-00208-y.
[17] Y. Kang, Z. Cai, C.-W. Tan, Q. Huang, H. Liu, Natural language processing (NLP) in management research: a literature review, J. Manag. Anal. 7 (2020) 139–172.
[18] M.J. Grant, A. Booth, A typology of reviews: an analysis of 14 review types and associated methodologies, Health Inf. Libr. J. 26 (2009) 91–108, https://doi.org/10.1111/j.1471-1842.2009.00848.x.
[19] D. Denyer, D. Tranfield, J.E. Van Aken, Developing design propositions through research synthesis, Organ. Stud. 29 (2008) 393–413, https://doi.org/10.1177/0170840607088020.
[20] W.R. King, Knowledge Management and Organizational Learning, in: Knowl. Manag. Organ. Learn., Springer, 2009, pp. 3–13.
[21] A. Tacchella, A. Napoletano, L. Pietronero, The language of innovation, PLoS One 15 (2020), e0230107.
[22] T.H. Davenport, L. Prusak, Working Knowledge: How Organizations Manage what They Know, Harvard Business Review Press, 2000.
[23] F. Ahmad, G. Widén, I. Huvila, The impact of workplace information literacy on organizational innovation: an empirical study, Int. J. Inf. Manag. 51 (2020), 102041.
[24] C. Muzzi, S. Albertini, Communities and managerial competencies supporting SMEs innovation networking: a longitudinal case study, R D Manag. 45 (2015) 196–211, https://doi.org/10.1111/radm.12060.
[25] E. Almirall, R. Casadesus-Masanell, Open versus closed innovation: a model of discovery and divergence, Acad. Manag. Rev. 35 (2010) 27–47, https://doi.org/10.5465/AMR.2010.45577790.

[26] B. Pikkemaat, M. Peters, C.-S. Chan, Needs, drivers and barriers of innovation: the case of an alpine community-model destination, Tourism Manag. Perspect. 25 (2018) 53–63, https://doi.org/10.1016/j.tmp.2017.11.004.
[27] S. Mancini, J.L.C. González, Role of technology transfer, innovation strategy and network: a conceptual model of innovation network to facilitate the internationalization process of SMEs, Technol. Invest. 12 (2021) 82–128, https://doi.org/10.4236/ti.2021.122006.
[28] E. Limaj, E.W.N. Bernroider, The roles of absorptive capacity and cultural balance for exploratory and exploitative innovation in SMEs, J. Bus. Res. 94 (2019) 137–153, https://doi.org/10.1016/j.jbusres.2017.10.052.
[29] M. Najda-Janoszka, S. Kopera, Exploring barriers to innovation in tourism industry – the case of southern region of Poland, Procedia - Soc. Behav. Sci. 110 (2014) 190–201, https://doi.org/10.1016/j.sbspro.2013.12.862.
[30] T. Kim, S.M. Lee, B. Lee, B. Lee, Seeking innovation: a typology of knowledge search strategy by SMEs, J. Strateg. Innov. Sustain. 15 (2020) 10–23.
[31] T. Leppäaho, K. Pajunen, Institutional distance and international networking, Enterpren. Reg. Dev. 30 (2018) 502–529, https://doi.org/10.1080/08985626.2017.1407365.
[32] M. Urbaniec, A. Żur, Business model innovation in corporate entrepreneurship: exploratory insights from corporate accelerators, Int. Enterpren. Manag. J. 17 (2021) 865–888, https://doi.org/10.1007/s11365-020-00646-1.
[33] S. Konsti-Laakso, T. Pihkala, S. Kraus, Facilitating SME innovation capability through business networking, Creativ. Innovat. Manag. 21 (2012) 93–105, https://doi.org/10.1111/j.1467-8691.2011.00623.x.
[34] W. Vanhaverbeke, M. Cloodt, Open Innovation in Value Networks, in: Open Innov. Res. New Paradigm, Oxford University Press, Oxford, 2006, pp. 258–281.
[35] J.J.P. Jansen, F.A.J. Van Den Bosch, H.W. Volberda, Exploratory innovation, exploitative innovation, and performance: effects of organizational antecedents and environmental moderators, Manag. Sci. 52 (2006) 1661–1674.
[36] M. Song, H. Park, K. Shin, Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean, Inf. Process. Manag. 56 (2019) 637–653, https://doi.org/10.1016/j.ipm.2018.12.005.
[37] J. Tissier, C. Gravier, A. Habrard, Dict2vec : Learning Word Embeddings Using Lexical Dictionaries, in: Proc. 2017 Conf. Empir. Methods Nat. Lang. Process, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 254–263, https://doi.org/10.18653/v1/D17-1024.
[38] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, M. Zhou, Sentiment embeddings with applications to sentiment analysis, IEEE Trans. Knowl. Data Eng. 28 (2016) 496–509, https://doi.org/10.1109/TKDE.2015.2489653.
[39] S. Xiong, H. Lv, W. Zhao, D. Ji, Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings, Neurocomputing 275 (2018) 2459–2466, https://doi.org/10.1016/j.neucom.2017.11.023.
[40] M. Zhang, B. Fan, N. Zhang, W. Wang, W. Fan, Mining product innovation ideas from online reviews, Inf. Process. Manag. 58 (2021), 102389, https://doi.org/10.1016/j.ipm.2020.102389.
[41] T. Roh, Y. Jeong, H. Jang, B. Yoon, Technology opportunity discovery by structuring user needs based on natural language processing and machine learning, PLoS One 14 (2019), e0223404, https://doi.org/10.1371/journal.pone.0223404.
[42] Z. Qiu, Z. Wang, Technology forecasting based on semantic and citation analysis of patents: a case of robotics domain, IEEE Trans. Eng. Manag. 69 (2022) 1216–1236, https://doi.org/10.1109/TEM.2020.2978849.
[43] S. Yun, W. Cho, C. Kim, S. Lee, nTechnological trend mining: identifying new technology opportunities using patent semantic analysis, Inf. Process. Manag. 59 (2022), 102993, https://doi.org/10.1016/j.ipm.2022.102993.
[44] J. Choudrie, S. Banerjee, K. Kotecha, R. Walambe, H. Karende, J. Ameta, Machine learning techniques and older adults processing of online information and misinformation: a covid 19 study, Comput. Hum. Behav. 119 (2021), 106716, https://doi.org/10.1016/j.chb.2021.106716.
[45] H. Wickham, Rvest: Easily Harvest (Scrape) Web Pages, 2022.
[46] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space, 2013, pp. 1–12. ArXiv. 1301.
[47] D. Selivanov, text2vec. http://text2vec.org/index.html, 2016 (accessed April 29, 2022).
[48] A. Roy, S. Pan, Incorporating extra knowledge to enhance word embedding, in: Proc. Twenty-Ninth Int. Jt. Conf. Artif. Intell., International Joint Conferences on Artificial Intelligence Organization, Yokohama, Japan, 2020, pp. 4929–4935, https://doi.org/10.24963/ijcai.2020/686.
[49] S. Kim, J. Zhang, Z. Chen, A. Oh, S. Liu, A hierarchical aspect-sentiment model for online reviews, Proc. AAAI Conf. Artif. Intell. (2013) 526–533.
[50] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: NIPS13 Proc. 26th Int. Conf. Neural Inf. Process. Syst., 2013, pp. 3111–3119, arXiv, http://arxiv.org/abs/1310.4546 (accessed July 12, 2022).
[51] J. Pennington, R. Socher, C.D. Manning, Glove: global vectors for word representation, in: Proc. 2014 Conf. Empir. Methods Nat. Lang. Process. EMNLP, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1532–1543, https://doi.org/10.3115/v1/D14-1162.
[52] R.J. Lennox, D. Veríssimo, W.M. Twardek, C.R. Davis, I. Jarić, Sentiment analysis as a measure of conservation culture in scientific literature, Conserv. Biol. 34 (2020) 462–471, https://doi.org/10.1111/cobi.13404.
[53] K. Taunk, S. De, S. Verma, A. Swetapadma, A brief review of nearest neighbor algorithm for learning and classification, in: 2019 Int. Conf. Intell. Comput. Control Syst. ICCS, 2019, pp. 1255–1260, https://doi.org/10.1109/ICCS45141.2019.9065747.
[54] D.K. Jain, P. Boyapati, J. Venkatesh, M. Prakash, An intelligent cognitive-inspired computing with big data analytics framework for sentiment analysis and classification, Inf. Process. Manag. 59 (2022), 102758, https://doi.org/10.1016/j.ipm.2021.102758.
[55] S. Neelakandan, M.A. Berlin, S. Tripathi, V.B. Devi, I. Bhardwaj, N. Arulkumar, IoT-based traffic prediction and traffic signal control system for smart city, Soft Comput. 25 (2021) 12241–12248, https://doi.org/10.1007/s00500-021-05896-x.
[56] N. Sharif-Razavian, A. Zollmann, An overview of nonparametric bayesian models and applications to natural language processing, Science (2008) 71–93.
[57] N. Nachar, The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution, Tutor. Quant. Methods Psychol. 4 (2008) 13–20.
[58] M.P. Fay, M.A. Proschan, Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules, Stat. Surv. 4 (2010) 1–39.
[59] F. Gralinski, A. Wróblewska, T. Stanisławek, K. Grabowski, T. Górecki, GEval: Tool for Debugging NLP Datasets and Models, 2019, pp. 254–262.
[60] A. Ly, Derks, How to Train a Machine Learning Model in JASP: Classification, JASP - Free User-Friendly Stat. Softw, 2019. https://jasp-stats.org/2019/10/07/how-to-train-a-machine-learning-model-in-jasp-classification/ (accessed April 25, 2022).
[61] F. Ali, D. Kwak, P. Khan, S. El-Sappagh, A. Ali, S. Ullah, K.H. Kim, K.-S. Kwak, Transportation sentiment analysis using word embedding and ontology-based topic modeling, Knowl.-Based Syst. 174 (2019) 27–42, https://doi.org/10.1016/j.knosys.2019.02.033.
[62] K. Pasupa, T.S.N. Ayutthaya, Thai sentiment analysis with deep learning techniques: a comparative study based on word embedding, POS-tag, and sentic features, Sustain. Cities Soc. 50 (2019), 101615, https://doi.org/10.1016/j.scs.2019.101615.
[63] B. Liang, H. Su, L. Gui, E. Cambria, R. Xu, Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks, Knowl.-Based Syst. 235 (2022), 107643, https://doi.org/10.1016/j.knosys.2021.107643.
[64] M.D. Mumford, K.S. Hester, I.C. Robledo, Chapter 1 - Creativity in Organizations: Importance and Approaches, in: M.D. Mumford (Ed.), Handb. Organ. Creat., Academic Press, San Diego, 2012, pp. 3–16, https://doi.org/10.1016/B978-0-12-374714-3.00001-X.
[65] F. Damanpour, D. Aravind, Chapter 19 - Organizational Structure and Innovation Revisited: from Organic to Ambidextrous Structure, in: M.D. Mumford (Ed.), Handb. Organ. Creat., Academic Press, San Diego, 2012, pp. 483–513, https://doi.org/10.1016/B978-0-12-374714-3.00019-7.
[66] G.I. Pereira, P.P. da Silva, D. Soule, Chapter 16 - Designing Markets for Innovative Electricity Services in the EU: the Roles of Policy, Technology, and Utility Capabilities, in: F. Sioshansi (Ed.), Consum. Prosumer Prosumager, Academic Press, 2019, pp. 355–382, https://doi.org/10.1016/B978-0-12-816835-6.00016-4.
[67] C. Gray, Absorptive capacity, knowledge management and innovation in entrepreneurial small firms, Int. J. Entrepreneurial Behav. Res. 12 (2006) 345–360, https://doi.org/10.1108/13552550610710144.
[68] S.A. Zahra, G. George, Absorptive capacity: a review, reconceptualization, and extension, Acad. Manag. Rev. 27 (2002) 185–203.

[69] M. Kesselring, S. Ruiz-Coupeau, M. Kirsch, F. Wagner, R. Gloaguen, Integrated test sites for innovation ecosystems, Technol. Innov. Manag. Rev. 11 (2021) 20–32, https://doi.org/10.22215/timreview/1454.
[70] D.Q. Nguyen, How firms accumulate knowledge to innovate – an empirical study of Vietnamese firms, Manag. Decis. 60 (2021) 1413–1437, https://doi.org/10.1108/MD-11-2020-1546.
[71] L. Cheng, T. Yu, A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems, Int. J. Energy Res. 43 (2019) 1928–1973.
[72] C. Burchardt, B. Maisch, Digitalization needs a cultural change–examples of applying Agility and Open Innovation to drive the digital transformation, Procedia Cirp 84 (2019) 112–117.