ARTICLE IN PRESS

Technovation xxx (xxxx) xxx

Contents lists available at ScienceDirect

Technovation

journal homepage: www.elsevier.com/locate/technovation

Exploring the boundaries of open innovation: Evidence from social media mining

Jose Ramon Saura [a,*], Daniel Palacios-Marqués [b], Domingo Ribeiro-Soriano [c]

[a] Rey Juan Carlos University, Madrid, Spain
[b] Universitat Politècnica de València, Valencia, Spain
[c] Universitat de València, Valencia, Spain

ARTICLE INFO

Keywords:
Open innovation
Social media mining
Data mining
User-generated content
Sentiment analysis

ABSTRACT

Technological development of the last several decades has driven open innovation towards organizational, business, social, and economic change. Open innovation has emerged as the main driver of change in a business sector that needs to be flexible and resilient, rapidly adapting to change through innovation. In this context, the present study aimed to explore the limits of open innovation by extracting evidence from user-generated content (UGC) on Twitter using social media mining. To this end, in terms of the methodology, we first applied machine learning Sentiment Analysis algorithm texted using Support Vector Classifier, Multinomial Naïve Bayes, Logistic Regression, and Random Forest Classifier to divide the sample of $n = 586.348$ tweets into three groups expressing the following three sentiments: positive, negative, and neutral. Then, we used a mathematical topic modeling algorithm known as Latent Dirichlet allocation to analyze the tweet databases. Finally, Python was used to develop textual analysis techniques under the theoretical framework of Computer-Aided Text Analysis and Natural Language Processing. The results revealed that, in the tweets dataset, there were eight topics. Of these topics, two contained tweets expressing negative sentiments (Culture and Business Models/Management), three topics contained tweets expressing positive sentiments (Communities, Creative projects and Ideas), and three topics contained tweets expressing neutral sentiments (Entrepreneurship, Teams and Technology). These topics are discussed in the context of limitations, risks, and characteristics of open innovation according to the UGC on Twitter. The paper concludes with the formulation of 20 limits of open innovation and 27 research questions for further research on open innovation, as well as a discussion of theoretical and practical implications of the study.

1. Introduction

In recent years, technology together with innovation processes has led to the emergence of new strategies for management (Dahlander and Gann, 2010; Belk, 2014), as well as promoted project development (Brunswicker and Chesbrough, 2018), community building (von Briel and Recker, 2017), branding (Wilson et al., 2015), use of social networks (Luqman et al., 2021), among other aspects (Huizingh et al., 2011).

In this paradigm, open innovation has been defined as a concept that encompasses novel challenges and practices of innovation processes (West et al., 2014; Bogers et al., 2018; Moretti and Biancardi, 2020). The concept of open innovation refers to management of innovation processes in corporations (Barham et al., 2020a), control over knowledge, as well as innovation-centered processes that take external ideas to drive positive and successful innovation (Ciesielska, 2018). Accordingly, open innovation proposes that, in their strategic innovation, companies should beyond their limits, seeking cooperation and partnerships with various organizations, professionals, and experts (Cavallo et al., 2021).

However, while open innovation has been extensively investigated in academic research, several authors argued that the limits of open innovation have to be studied and discussed in terms of risk (Jugend et al., 2020; Hervas-Oliver et al., 2021; Marshall et al., 2021), processes and practices (Pilav-Velic and Jahic, 2021), management (Bagherzadeh et al., 2019), information (Adamides and Karacapilidis (2020)), finance (García-Quevedo et al., 2018), or technology (Noh and Lee, 2020).

The development of the concept of open innovation in the last

; UGC, User-Generated Content; UGD, User-Generated Data; CATA, Computer-Aided Text Analysis; NLP, Natural Language Processing; LDA, Latent Dirichlet Allocation; SVC, Support Vector Classifier; MNB, Multinomial Naïve Bayes; LR, Logistic Regression; RFC, Random Forest Classifier.

* Corresponding author.
E-mail addresses: joseramon.saura@urjc.es (J.R. Saura), dapamar@doe.upv.es (D. Palacios-Marqués), domingo.ribeiro@uv.es (D. Ribeiro-Soriano).

https://doi.org/10.1016/j.technovation.2021.102447
Received 19 October 2021; Received in revised form 12 December 2021; Accepted 19 December 2021
Available online 4 January 2022

several decades has occurred in parallel to the exponential growth of the use of social networks (Saura et al., 2021). Technological development and innovation have led to the emergence of new social networks that meet user needs (Liu et al., 2021). In a connected digital ecosystem where huge amounts of data are generated on a daily basis, social networks have been configured as optimal systems to explore the world (Ribeiro-Navarrete et al., 2021), create knowledge, and extract insights on both traditional and novel topics (Tandon et al., 2021).

In this context, massive amounts of data known as user-generated content (UGC) have come to be considered as a valid source of information. UGC is extracted from social networks to identify insights and create knowledge through evidence collected from social media mining (Krumm et al., 2008).

In this paradigm, it is imperative to define the boundaries, characteristics, and benefits of open innovation and the use of UGC used to extract insights from large amounts of data (Krippendorff, 2018). In response to this need, the present study explores UGC from the social network Twitter to define the limits of open innovation through the application of sentiment analysis and topic modeling algorithms that work with machine learning.

The novelty of the present studies lies in the fact that, owing to the data-mining technologies used, and filling the gap in available research, we identify and discuss the boundaries of open innovation evidenced by UGC.

The research questions addressed in the present study are as follows: *RQ1: What are the limits of open innovation according to UGC on Twitter? RQ2: What is the sentiment of the topics that characterize open innovation according to UGC on Twitter?* To answer these two questions, in the present study, we aim to achieve the following objectives:

- To create knowledge about open innovation
- To Identify new analytical perspectives for the limits of open innovation
- To explore the characteristics and sentiments of open innovation according to UGC from Twitter
- To outline future guidelines for open innovation research using social networks
- To establish guidelines for the use of information classification models in social networks

Methodologically, in the present study, we followed the approaches previously proposed by Hiremath and Patil (2020), Ozcan et al. (2021) and Saura et al. (2021a). Specifically, for data analysis, we first applied Textblob Sentiment Analysis using Support Vector Classifier (SVC), Multinomial Naïve Bayes (MNB), Logistic Regression (LR), and Random Forest Classifier (RFC). These approximations were performed to divide the sample of 586,348 tweets into three subgroups expressing the following three sentiments: positive, negative and neutral. In the next step, a mathematical thematic modeling algorithm known as Latent Dirichlet allocation (LDA) was used to analyze the tweet databases. The results were validated with the *p*-value and keyness indicators. Finally, Python was used to develop textual analysis techniques under the theoretical framework of Computer-Aided Text Analysis (CATA) (Leon, 2009) and Natural Language Processing (NLP) (Hirschberg and Manning, 2015) through computing the count and weighted percentage variables.

The remainder of this paper is structured as follows. In Section 2, we review relevant literature. Section 3 presents the methodology. The results are reported in Section 4. The results and discussed and a roadmap for future research on open innovation are provided in Section 5. Finally, conclusions, theoretical and practical implications, as well as limitations of the present are identified in Section 6.

Table 1
Relevant previous studies on innovation development.

Authors	Description
Leonardi (2014)	Emphasized that innovative social media strategies drive knowledge generation, improve communication, and enable new digital scenarios for creative innovation.
Nguyen et al. (2015)	Highlighted brand innovation and knowledge acquisition through social media and strategic capabilities
Roberts and Piller (2016)	Defined and identified the roles that companies should take to drive innovation both in social media and in the sectors in which companies operate
Muninger et al. (2019)	Investigated how companies should drive innovation through the use of new digital channels such as social media. Defined the capabilities that allow companies to innovate.
Fischer et al. (2021)	Analyzed the main means for radical innovation in the digital age and identified the characteristics of innovation in the digital age and how companies should drive change

2. Literature review

2.1. Innovation, open innovation and social media mining

Open innovation models are applied by companies to develop successful innovation strategies (Zhu et al., 2019). However, in the processes of development, construction and implementation of new ideas (Gatzweiler et al., 2017), in order to accelerate the innovation of business activities, firms require external support (West and Bogers, 2014), critical knowledge, as well as the identification of new ways to obtain and/or create knowledge (Bogers et al., 2017).

In available literature, there is controversy over the main limitation of open innovation and the causes or drivers of success. In this respect, Vanhaverbeke et al. (2017) demonstrated that there are organizational barriers to propose correct open innovation models and overcome cultural barriers and legal regulations that may affect these processes. In addition, Greco et al. (2018) argued that financial and economic drivers, or conflicting demands in the development of new ideas, can also become barriers to innovation.

While open innovation is admittedly a very complex process, it is central to companies' innovation strategies. The flows of information, technologies, and ideas are all building blocks of the innovation process (Alassaf et al., 2020). In this connection, Fischer et al. (2021) argued that social networks can serve as a relevant source of knowledge creation to drive change in organizations. Accordingly, social networks could also help to improve open innovation processes. As argued by Chesbrough and Bogers (2014), to become truly innovative, firms should be open and willing to accept external knowledge. Yet, before the application and use of innovative technologies, their challenges should be discussed (Mention, 2011).

For instance, while Lifshitz-Assaf et al. (2018) argued that open innovation should be based on the development of open innovation projects that integrate external knowledge through social networks and digital platforms, Barham et al. (2020b) noted that there are risks concerning open innovation when there is no internal support that defines the company's commitment to open innovation. Accordingly, Barham et al. (2020b) underscored the importance of the role of managers and their strategic support for open innovation. Similarly, West et al. (2014) also indicated that there are negative attitudes towards external knowledge and cooperation through communication in companies.

Notwithstanding, Leonardi (2014) indicated that external knowledge acquired in digital ecosystems can enhance creative innovation and boost knowledge generation. Similarly, Enkel et al. (2009) also indicated that strategic orientation is a fundamental pillar for open innovation to be successfully integrated into business processes.

Following these considerations, Abouzeedan and Hedner (2012) highlighted the importance of internal weaknesses in companies to make their processes more flexible and to implement the open innovation projects. Both from the technological, organizational, and project

decision-making perspective, companies do not make sufficient effort to establish additional routines that drive open innovation as a real possibility for their success (Chesbrough and Crowther, 2006).

To provide a better understanding of how open innovation can be studied not only from theoretical perspectives and frameworks, but also from a practical perspective where social networks and digital ecosystems are used as a primary data source, Table 1 summarizes previous studies on innovation development. All studies reviewed in Table justified the creation of knowledge and improvement of innovation through the study of digital data sources.

As mentioned above, although the adoption of open innovation strategies has been extensively studied in the literature in the last decade, previous approaches focused on the digital environment play am increasingly more relevant role in the development of the concept of open innovation. As can be seen in Table 1, digital ecosystems, knowledge, and insights extraction in social networks or opportunities of technological developments and business models on the Internet drive researchers to take digital data sources such as a fundamental source for the development and understanding of the concept of open innovation (Elia et al., 2020).

In this connection, the concept of social media mining (Di Domenico et al., 2021) is directly linked to the UGC and consists of collecting large amounts of data from social media profiles to identify patterns and insights in relation to a specific topic. In the present study, the use of UGC extracted from Twitter is used as the source of information on which knowledge about open innovation is created (see also Klein et al., 2021; Saura et al., 2021a).

2.2. User-generated content (UGC)

In social sciences, UGC is conventionally used as a source of valid information extracted from content posted by users on social networks (Luca et al., 2015). UGC includes small pieces of text that express different opinions and feelings linked to a specific topic (North-Kim and Johnson, 2016). UGC can include comments, reviews, posts, tweets, videos comments, or any other source of information that users generate in digital environments and social networks.

In the present study, the UGC is used to understand the characteristics of open innovation, as it can be used to understand how a specific industry works. For instance, Roma and Aloini (2019) demonstrated that UGC can be meaningfully used to understand the opinion of users in relation to brands. Accordingly, UGC has been studied to better understand the society (Östman and Chang, 2012), data privacy (Ribeiro-Navarrete et al., 2021), strategies focused on innovative decision-making (Saura et al., 2021a), or economics (Dhar and Chang, 2009), among many other topics.

There is a broad consensus among social science scholars that the content published by social network users can provide new perspectives on the emerging issues or link the already available knowledge about an industry to new academic contributions. At the same time, in recent years, social networks and traditional/digital advertising (Minowa and Belk, 2017) have become a valid channel for research and a source of data for social innovation (Pelka and Kaletka (2011), online user behavior (Van Dijck, 2009), or event prediction (Yang et al., 2014).

Similarly, public opinion—investigated through the analysis of social media content and UGC (Mehmet and Clarke, 2016)—can help to identify and operationalize new variables and indicators that can further be quantitatively explored in future research (Leung, 2009), as well as used to clarify emerging paradigms or establish new metrics and practices in relation to how companies carry out professional activities.

Overall, due to the application of technologies that work with machine learning, the use of social media listening techniques and innovation has been argued to enable the creation of value (Chesbrough et al., 2018), collection of insights (Shan et al., 2020), as well as extraction of knowledge. Accordingly, Grover et al. (2019) concluded that UGC is one of the applications that can be adopted to clarify emerging paradigms and identify new patterns concerning the object of study (Daim et al., 2016). Therefore, the UGC is consensually used in the literature as one of the main sources of data for the development of studies based on social media listening. Viewed from this perspective, relevant previous studies that used UGC to create knowledge and improve understanding of sectors related to innovation, its development and application are summarized in Table 2.

Table 2
Relevant previous studies on user-generated content and knowledge creation.

Authors	Description
Lexhagen et al. (2012)	Analyzed social media content in the form of UGC to extract insights and create knowledge in relation to innovation in tourism destinations, management, and marketing.
Schmunk et al. (2013)	Extracted UGC to improve decision-making and knowledge creation processes on strategic and technological levels.
Zhang et al. (2018)	Investigated the impact of UGC opinions on social networks to measure their influence on innovation and change processes.
Hill et al. (2019)	Measured brand audiences in relation to demographic interests through the analysis of UGC in online chats.
Ho-Dac (2020)	Defined the characteristics and value of UGC for the improvement of products and services and described the value of UGC in driving strategy development and social media listening.
Daradkeh (2021)	Explored the use of UGC for Business Intelligence in Innovation; extracted UGC information from an Open Innovation Community platform.

3. Methodology

The methodology used in the present study is based on social media-mining techniques divided into three distinct phases supported by the CATA theoretical framework (Short et al., 2010). CATA defines the different approaches that can be developed with computer-aided text analysis techniques (Park et al., 2019). These techniques can be meaningfully used to build statistical validation, propose theoretically-based constructs and variables, or create knowledge by identifying insights to be studied in future research (Täuscher et al., 2020).

As argued by Xuanyang et al. (2005) and Leon (2009), in order to increase validation capacity of results in CATA-centered approaches, technologies working with machine learning or mathematical algorithms that drive prediction can be used. Furthermore, another meaningful methodology for knowledge creation using CATA through the identification of insights is content analysis (Krippendorff, 2018).

In the present study, the data were analyzed in the following three steps. First, we applied sentiment analysis with Textbob using SVC, MNB, LR, and RFC experiments to obtain the highest accuracy in the results. Second, a LDA was developed to divide the database into corpora of tweets expressing negative, positive, or neutral sentiment. Thirdly and finally, through textual and content analysis using Python (Bhavsar and Manglani, 2019), insights were extracted from the sample.

3.1. Data sampling

A sample of tweets (n = 586,348 tweets) was extracted from the Twitter API. Twitter queries (hashtags) used for data extraction were as follows: #Innovation AND #OpenInnovation or #OpenInnovation. Data collection took place between May 2021 and October 2021. Of note, the data extraction was carried out during the time period when there was no event linked to the open innovation industry that could modify the sample following a non-probability sampling frame (Lehdonvirta et al., 2021). Non-probability sampling (Sihombing, 2017) is an approach to sample elaboration, which is also known as judgment sample. In the present study, we followed Lehdonvirta et al. (2021) criteria and knowledge of the object of study to develop the sampling process considering non-probabilistic variables. Non-probability sampling is commonly used in studies that work with UGC in social networks

(Bahtar and Muda, 2016; Saura et al., 2021a).

The study database was filtered using Python and Pandas libraries with the following criteria. First, repeated tweets were removed to obtain as many original tweets as possible. Retweets were also considered duplicate content, because they express content previously published in the social network. URLs and links were removed, and only text was analyzed using Natural Language Processing (NLP) techniques. All included tweets contained a minimum of 80 characters (Kim et al., 2018). Images and videos contained in the tweets were not analyzed because they were not the subject of our study (Palguna et al., 2015). After all filtering steps mentioned above, a total of 408,802 tweets were retained in the dataset.

3.2. Textblob sentiment analysis

In general, sentiment analysis algorithms are used to divide databases into different categories or classifications. Many algorithms can be trained with machine learning processes to improve the accuracy of results. The databases on which sentiment analysis algorithms work can be of a diverse nature. However, as argued by Sibona et al. (2020), user-generated data (UGD) or UGC is a valid source for the use and application of this data-mining technique.

Using Textblob (Hiremath and Patil, 2020) and supported by the CATA and NLP framework. We first used 1581 tweets to train the Textblob algorithm to divide the data into tweets expressing three types of sentiment: positive, negative, and neutral. Sentiments expressed in emoticons, images, or videos were not considered (Zafar et al., 2015). The algorithm training process was performed manually. To this end, the dataset of 1581 tweets was divided into equal parts and classified without comparing the results, adding values of 1 for positive tweets, 0 for neutral tweets, and −1 for negative tweets.

Once the samples were classified, in order to evaluate the accuracy of the algorithm, for a sentiment analysis using Textblob, polarity scores were measured and classified as polarity or subjectivity. Polarity ranged from −1 to 1, while subjectivity ranged from 0.0 to 1.0, where 0.0 is very objective and 1.0 is very subjective. This accuracy measure commonly used in the classification for the results of the models that work with sentiment analysis.

Likewise, Textblob is a well-known library developed in Python that is used to find common text processing operations. Textblob is developed in NLTK and patterns. Of note, one of the limitations of algorithms that work with sentiment is that these algorithms do not recognize irony, sarcasm, and connotations. In the process of training the algorithm, we considered this limitation in an attempt to increase the quality of the results (Hasan et al., 2018).

Finally, to validate the results of our sentiment analysis using Textblob, following Hiremath and Patil (2020), we computed 4 cross--validations—namely, taking into account the variables of precision, recall, f1-score and support. These variables were supported in the results in terms of macro average and weighted average. Therefore, for the classification of patterns in the analyzed database, the following technologies were used for the computation of the Textblob model:

- Support Vector Classifier (Lau and Wu, 2003)
- Multinomial Naïve Bayes (Jiang et al., 2016)
- Logistic Regression (Kumar et al., 2016)
- Random Forest Classifier (Al Amrani et al., 2018)

3.3. Latent Dirichlet allocation (LDA) model

To subdivide the database into topics, the mathematical algorithm known as LDA was used. We chose LDA as a unsupervised machine learning algorithm due to its flexibility and adaptability for the analysis of databases consisting of tweets. However, to perform this process, other valid approaches are Latent Semantic Analysis (LSA) or Non-Negative Matrix Factorization. Following Jelodar et al. (2019), the LDA algorithm was developed in Python LDA 1.0.5 using Gibbs sampling (MAC version).

As mentioned above, based on the results of sentiment analysis, the main database was divided into three databases with tweets expressing three different sentiments. To these databases, LDA was applied to try to find a probability of hidden distributions in the input data.

In general, the LDA algorithm allows for the identification of topics in commonly structured as observations (documents) composed by tweets. However, in the present study, we dealt with data sources—namely, Twitter-based UGC structured as small pieces of text. Applying LDA to databases divided into sentiments, we identified topics linked to these three types of sentiments (Resnik et al., 2015).

Consequently, LDA was configured as a topic-modeling technique classified within data mining, and we studied different inputs in the words contained in the databases. In this way, the model was used to identify the most frequent words and thus to name those topics (Zhao et al., 2011).

Therefore, a probabilistic assumption developed in the LDA was used to first identify the distribution of the themes in the analyzed database and then automatically group the words that made up the themes that were subsequently labelled using the words frequently occurring in the data (Tajbakhsh and Bagherzadeh, 2019). Once these data were obtained, we analyzed the groups of words obtained from the analyzed databases to select topics relevant for our research objective (see the Results section for further detail).

3.4. Textual and content analysis

Upon obtaining the results of the sentiment analysis and LDA, we applied textual and content analysis following NLP. Specifically, variables such as weighted percentages, count and keywords frequency were computed to extract insight from the data (Klein et al., 2021).

Overall, textual analysis makes it possible to understand the weight of each of the words in the database, which facilitates identification of patterns in the data. This type of exploratory approach allows for the construction of insights and identification of drivers that explain the object of study of the research. Moreover, additional approaches such as n-gram analysis can also be developed (Linvill et al., 2012).

In the present study, following Saura et al. (2021a) who used Latin numerical prefixes, an n-gram of size 1 was referred to as a "unigram"; size 2 was a "bigram". Therefore, in the process of textual analysis, n-grams relative to unigram (positive, negative, and neutral) as well as bigram (negative, positive, and neutral) were identified. These processes were also previously used by Utama (2019) and Widaretna et al. (2021) using the concept of Mutual Information (MI), which validates whether there are correlations between words and indicators compared in NLP and data-mining approaches.

4. Results

4.1. Results of sentiment analysis

The results of our sentiment analysis with four standard classifier methods (FRC, SVC, MNB, and LR) were first considered in relation to the obtained level of accuracy. As mentioned above, the accuracy variable measures the percentage of success in the results of a model that has been previously trained. When working with machine learning, accuracy is one of the most relevant variables (Dhaoui et al., 2017).

Among the five cross-validation experiments that we performed (Sl. No. from 0 to 4), the ones that obtained the highest accuracy were Linear SVC Sl. No. 3 (0.902100) and 1 (0.884401). The remaining classifier methods obtained accuracy for Sl. No. 3 of 0.693041 in RFC, Sl. No. 4 of 0.762001 in MNB and, finally, Sl. No.1 of 0.841202 in LR.

Within the standard classification development processes in sentiment analysis, multiple validation scales are a standard procedure to give robustness to the applied models in terms of accuracy (Li and Liu,

Table 3
Model category details.

Sl. No.	Model Name	Fold_idx	Accuracy - Textblob
0	RandomForestClassifier	0	0.650310
1	RandomForestClassifier	1	0.670420
2	RandomForestClassifier	2	0.668031
3	RandomForestClassifier	**3**	**0.693041**
4	RandomForestClassifier	4	0.691401
5	LinearSVC	0	0.850574
6	LinearSVC	1	0.863718
7	LinearSVC	2	0.884401
8	LinearSVC	**3**	**0.902100**
9	LinearSVC	4	0.870216
10	Multinomial Naïve Bayes	0	0.742092
11	Multinomial Naïve Bayes	1	0.732970
12	Multinomial Naïve Bayes	2	0.743102
13	Multinomial Naïve Bayes	3	0.710522
14	Multinomial Naïve Bayes	**4**	**0.762001**
15	LogisticRegression	0	0.811002
16	LogisticRegression	**1**	**0.841202**
17	LogisticRegression	2	0.813120
18	LogisticRegression	3	0.824021
19	LogisticRegression	4	0.840431

Source: the authors

Table 4
Summarized brief scores.

Sl. No.	Model Name	Scores of Textblob analysis
1	LinearSVC	**0.902100**
2	LogisticRegression	0.841202
3	MultinominalNB	0.762001
4	RandomForestClassifier	0.693041

Source: the authors

2014). Table 3 shows the results for each of the experiments developed (total n = 19); a number from 0 to 4 was applied to the classification scale for each model used and the corresponding accuracy.

Similarly, in order summarize the approximation and the highest accuracy values, the highest accuracy values according to the classification model used and the result are shown in Table 4. The highest accuracy result for SVC was 0.902100; 0.841202 for LR; 0.762001 for MNB, finally, 0.693041 for RFC.

Following Ozcan et al. (2021) who highlighted the relevance of data visualization in machine learning-centric approaches, Fig. 1 (a) shows the results of experiments with 5 cross-validations results (see also Hiremath and Patil, 2020). Fig. 1 reflects a maximum total accuracy of 0.99 on the X-axis. On the Y-axis, the numbers related to the experiments performed with LSVC, RFC, LG, and MNB with a total of 19 are presented. The highest accuracy achieved according to the classification model used is highlighted in bold.

Similarly, in the analysis processes using sentiment analysis, it is a standard procedure to represent the variables obtained concerning the accuracy, recall, f1-score and support with regard to different sentiments (in our study—positive, negative, and neutral).

In this context, accuracy is an indicator that measures the quality of the machine learning model for the development of the tasks assigned to it. Furthermore, the recall variable reflects the amount of the machine learning model used that can be identified in a database. F1-score, which is a metric that combines accuracy and recall into a single indicator, is a practical indicator because that facilitates the comparison of performance in terms of accuracy and completeness analysis between various solutions (Saura et al., 2021). Finally, the support variable evaluates how much support of machine learning is used to make model predictions—i.e., how much the value of the learning machine has been used to produce the prediction on which the accuracy is measured.

The values of accuracy, recall, f1-score and support are summarized in Table 5, in which the macro average variable measures the average

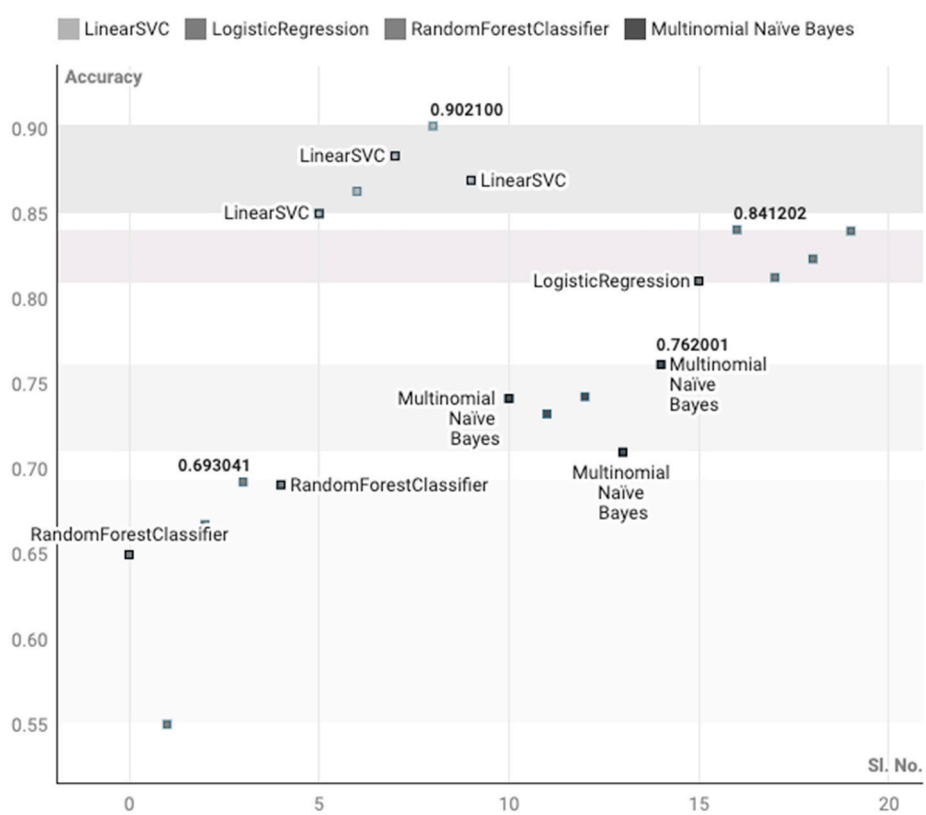

Fig. 1. Main brief scores of TextBlob Analysis in 5 cross-validation experiments
* Sl. No. The number of programmed experiments corresponding to the series of 5 cross-validation results. Source: the authors.

Table 5
Sentiment analysis classification report.

Sl. No.	Parameters	Vader			
		precision	recall	f1-score	support
1	Negative	0.79	0.89	0.77	20.012
2	Positive	0.88	0.77	0.83	23.429
3	Neutral	0.90	0.96	0.93	20.371
4	Accuracy	–	–	0.88	43.471
5	Macro avg	0.85	0.86	0.76	43.471
6	Weighted avg	0.83	0.84	0.87	43.471

Source: the authors

Table 6
LDA topics results.

R	Topics	Description	Sent.*	Keyness	*p*-value
1	Culture	Obsolete and inflexible business cultures that do not support open innovation with commitment and are seen as barriers	Ne	702.91	0.039
2	Communities	Building collaborative communities motivated about innovation	P	547.56	0.031
3	Creative projects	Creative projects focused on open innovation	P	538.85	0.030
4	Entrepreneurship	Entrepreneurship, as a precursor to the development of open innovation, is based on technological and innovative development	N	402.01	0.028
5	Ideas	To find the support from companies to employees' ideas in terms of action, funding, and structural support	P	396.30	0.026
6	Teams	Effective team building to drive open innovation in companies	N	205.84	0.014
7	Technology	The fundamental role of technology in open innovation processes	N	201.07	0.013
8	Business Models & Management	Depending on the type of business model, open innovation and its management determine business success	Ne	174.25	0.010

* Sent. = Sentiment: N = Neutral sentiment; Ne = Negative sentiment; P = Positive sentiment.
Source: the authors

total of the model based on the variables analyzed, while the weighted average measures its relativity in terms of weight.

Of note, the highest value for the recall variable was obtained for neutral and negative tweets, with the corresponding values of 0.89 and 0.96, respectively. Positive tweets obtained a 0.77 recall.

4.2. LDA results

Based on the results of sentiment analysis, we identified a total of 8 topics about open innovation. Of note, however, the LDA results sometimes have to be adapted to match the research objectives. Accordingly, we analyzed the first 20 topics identified for each database analyzed by the LDA model.

Of the 20 topics identified for each database according to their sentiments, we selected 3 positive topics, 3 neutral topics and 2 negative topics. Them using the first 10–15 most frequent words in each of the topics, we labelled the topics. Topic sentiments were directly linked to the previous step of the methodology, in which the sentiment analysis was used to divide the dataset into three, each one according to one sentiment: positive, negative and neutral.

The descriptions of the topics were made based on the identification of the subject and the words that compose it, taking into account their context and the objectives of the research. Although this selection was carried out manually, it is a standardized process in LDA approximations (Jelodar et al., 2019). Table 6 presents the topics with their descriptions and corresponding sentiments.

In addition, in order to establish the relevance of the different topics identified, the value of the keyness variable was calculated. This variable measures the strength of the link between the topics. In statistical terms, the keyness variable determines the log-likelihood score or chi-square statistics values (Rayson and Garside, 2000). Therefore, keyness evaluates the extent of the frequency difference and is used to study the corpus of words that make up the topics. These processes were also carried out in previous studies that used mutual information and discourse analysis (Guizzo, 2019) in which the frequencies of the words selected within each topic were compared to the total size of the sample. Therefore, following Gabrielatos (2017), to measure the relevance of each topic in mutual information and discourse analysis, log-likelihood of >3.8 was statistically significant when the p-value < 0.05. Although the results were not statistically significant, they can be used to understand the relevance of each theme by analyzing the results of the p-value and keyness variables (Pojanapunya and Todd, 2018) compared to the rest of the dataset. The corresponding results for keyness, p-value, sentiment, as well as topics names and description are summarized in Table 6.

In addition, in order to visually represent the identified topics in relation to their relevance in terms of keyness in the analyzed database, Fig. 2 shows a map of the topics according to their relationship to the center (i.e., open innovation).

4.3. Textual analysis results

Furthermore, to obtain additional insights, the number of times that the keywords were repeated in the databases as analyzed in terms of frequency, and weighted percentage (WP) in the total database (Loughran and McDonald, 2016). WP is an indicator that presents, in terms of percentages, the weight of a node (set of words) with respect to the total number of words. For the calculation of this value, similar keywords were grouped into nodes or groups of words to calculate their weight. This value can be identified with Python or software such as NVivo (Hirschberg and Manning, 2015).

To this end, we used the standard approaches in NLP and CATA and the indications of Krippendorff (2018). Pandas GroupBy in Python was used for data filtering and analysis. Frequency was determined by text nodes grouped together and containing the total number of words that made them up.

In this way, even if an idea was expressed with different words, its analysis was categorized within the node of a specific topic, thus increasing the presence of the word in the databases in relation to the topic represented by that node (Krippendorft, 2018). Table 7 presents the results of our analysis of frequency and WP. Some of the words in Table 7 exemplify the content of each topic.

Subsequently, in order to extract additional insights, we performed n-gram analysis. Overall, CATA-based studies can be complemented with information extracted using corpus linguistic methods. In our case, an n-gram predicted the frequency of a word based on the occurrence of its n-1 previous word. Similarly, a bigram model ($n = 2$) predicted the occurrence of a word given only its previous word (as n-1 = 1).

Table 8 presents a list of n-grams (unigrams and bigrams) and their frequencies in relation to the keywords or collocates. Such insights make it possible to consider the context of words and their relevance in a corpus (McEnery and Hardie, 2013).

This context is defined here as words used together and in which a

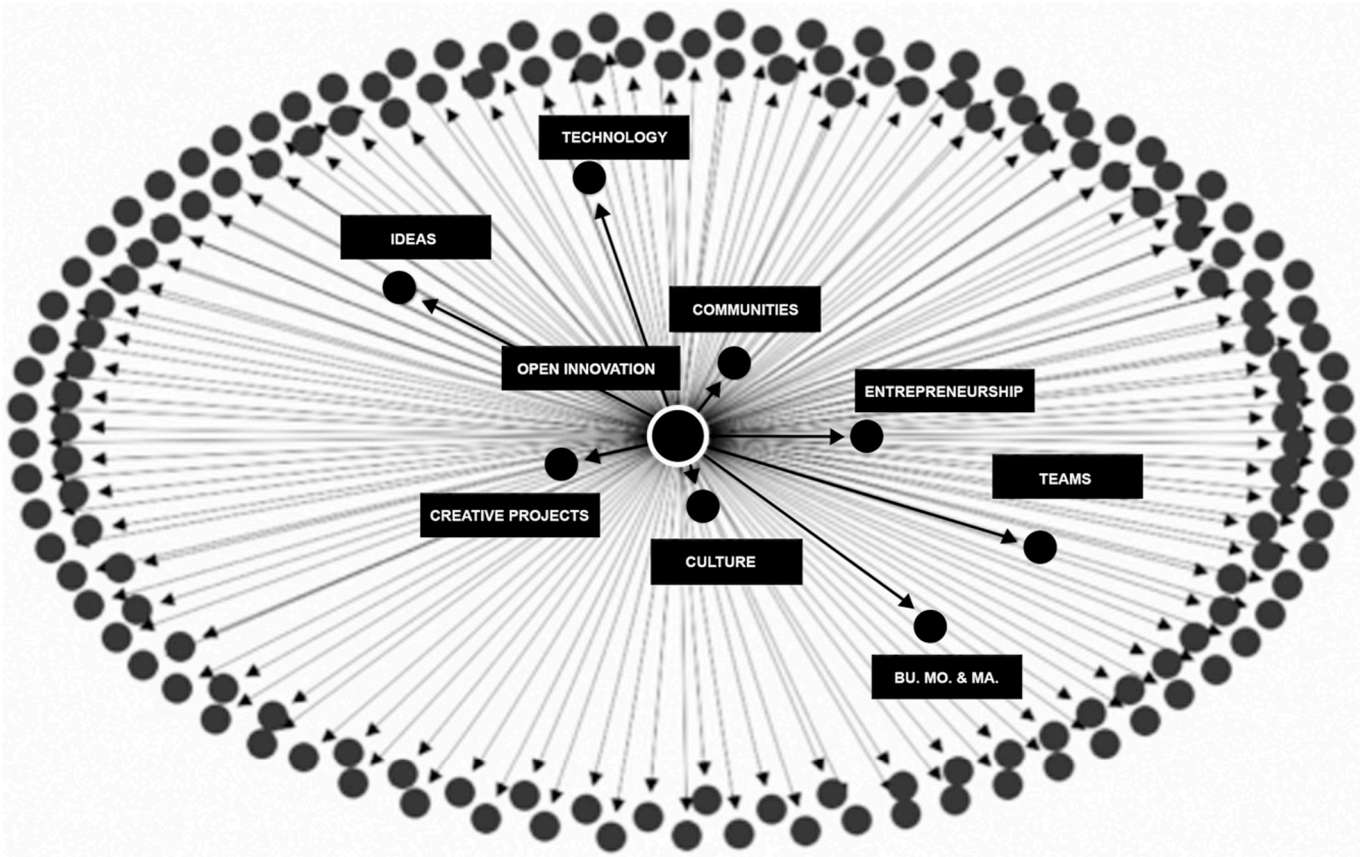

Fig. 2. Visual representation of identified topics in relation to open innovation.
Source: the authors.

Table 7
Grouped keywords, count, and weighted percentage.

R	Topics	Similar words	Frequency	WP
1	Culture	cultures, internal culture, culture transformation, culture development, etc.	102,031	17.36
2	Communities	innovative communities, networks, community innovations, collective intelligence, community members, etc.	93,793	15.02
3	Creative projects	projects ideation, sharing ideas, collaborative ideas, project management, projects risks, etc.	91,052	14.97
4	Entrepreneurship	entrepreneurial orientation, entrepreneurship platforms, entrepreneurs ideas, open innovation and entrepreneurship, etc.	64,936	10.39
5	Ideas	open innovation ideas, open innovation examples, open innovation best ideas, open innovation solutions, etc.	60,614	9.74
6	Teams	open innovation teams, experts, departments, training, motivated teams, etc.	29,080	4.60
7	Technology	technologies developments, technological tools, technological resources, advances in technology, etc.	23,894	3.75
8	Business Models & Management	traditional business models, senior management, vertical structures, etc.	12,583	2.02

Source: the authors

lexical relationship or n-gram package is established. In this sense, these words can provide additional information and can be used to complement the results. The collocates for the selected themes, Frq. and Frq. L and Frq. R, are presented in Table 8.

5. Discussion

In recent years, open innovation has emerged as a driving force for many companies and industries. However, as argued by Abouzeedan and Hedner (2012), there are still limitations and barriers to developing the appropriate corporate culture to successfully support open innovation (Dabić et al., 2019).

Accordingly, in our results, corporate culture for open innovation was found to be a negative topic in our data (keyness 702.91 and *p*-value 0.039). At present, there are many obsolete and inflexible company cultures that do not provide sufficient strength and commitment for employees to feel that open innovation is a relevant part of their organization. As argued by Huizingh et al. (2011), internal company corporate culture, strategic transformation models, and the development of a culture that supports open innovation are key factors to ensure that there are no limits to open innovation in companies (Zenobia et al., 2009).

Furthermore, von Briel and Recker (2017) noted that collaborative communities boost innovation and new ideas. In line with this observation, a positive topic that we identified in our analysis was communities (keyness 547.56 and *p*-value 0.031); this topic suggests that the construction of collaborative communities keeps people motivated and drives the development of ideas focused on open innovation. Also, Fichter (2009) argued that the very open nature of innovation-focused communities enables new opportunities to develop open innovation. Similarly, Martínez-Torres (2014) noted that the feeling of belonging to

Table 8
N-grams for the collocates of selected topics.

*R	Collocates for the topic "Culture"				R	Collocates for the topic "Communities"			
	*Freq	*Freq L	*Freq R	Topic		Freq	Freq L	Freq R	Topic
1	69261	32081	37180	Culture	1	40127	17592	22535	communities
2	16981	8445	8536	Development	2	14803	13021	1782	innovative
3	4205	2180	2025	innovation culture	3	8901	4691	4210	members
4	4053	1984	2069	Companies	4	3585	1802	1783	intelligence
5	1047	793	254	Organizational	5	1301	673	628	users
R	Collocates for the topic "Creative projects"				R	Collocates for the topic "Entrepreneurship"			
	Freq	Freq L	Freq R	Topic		Freq	Freq L	Freq R	Topic
1	30941	5740	25201	Projects	1	23756	10782	12974	entrepreneur
2	14712	7291	7421	project management	2	16930	6092	10838	entrepreneurship
3	6009	5780	229	Creative	3	4901	2469	2432	orientation
4	3811	1063	2748	Ideas	4	2301	1360	941	platform
5	748	117	631	Risks	5	1830	381	1449	thinking

*Rank (R) *Frequency (F) *(F) of words on the left (FreqL) *(F) on the right (FreqR).

a community increases the ability of both employees and members of that community to feel supported and motivated in relation to the development, execution, and implementation of ideas focused on open innovation.

In line with previous research that concluded that creative projects are the fundamental axis for open innovation (Du et al., 2014), in our results, we found a positive topic related to creative projects (keyness 538.85 and *p*-value 0.030). As argued by Boscherini et al. (2010), creative projects focused on open innovation promote idea sharing, collaborative construction of new ideas, creative thinking, and effective identification of risks. Similarly, Füller et al. (2012) showed that creative projects should be supported by company management, as the creative nature of ideas that emerge in such projects drives the development of open innovation.

Furthermore, in line with Schweitzer et al. (2021) conclusion that entrepreneurship is directly linked to innovative, technological, and open innovation processes, we identified a neutral topic with entrepreneurship (keyness 402.01 and *p*-value 0.028) and open innovation. This topic suggests that there are important links between entrepreneurial projects and open innovation through the use of digital platforms or technologies (Del Vecchio et al., 2018; Saura et al., 2021b), open tools and collaborative activities that base their philosophy and culture on open innovation as a central axis.

Next, Bughin et al. (2008) developed new models to understand how new ideas should be constructed, while Hofstetter et al. (2021) highlighted the importance of new idea development processes in companies to drive change and increase profitability. In agreement with the above, in our results, we found a positive topic of ideas (keyness 396.30 and *p*-value 0.026) in terms of action, funding and structural support. Therefore, it can be concluded that employees positively perceive companies' support and implementation of their ideas concerning open innovation.

Furthermore, authors such as Igartua et al. (2010) indicated that the internal management of ideas in companies is essential for employees to feel motivated. Accordingly, the construction of new open innovation ideas by acquiring knowledge external to the organizations can cause frustrations for the employees of these companies.

Teams have always been a fundamental part of open innovation. Similarly to Chatenier et al. (2010) finding that multidisciplinary and participative teams must be developed to obtain effective open innovation processes, in our data, we identified the neutral topic of teams (keyness 205.84 and *p*-value 0.014). This topic suggests that teams should be composed of experts from different departments of the company to increase multidisciplinary ideas in a creative project (see also Markovic et al., 2020).

Next, with regard to technology and the development of open innovation, Bigliardi et al. (2021) showed that technological advances allow companies to develop open innovation. Accordingly, in our results, we identified a neutral topic in relation to technology (keyness 201.07 and *p*-value 0.013) and open innovation as discussed by Fortunato et al. (2017). It is clear that technology is a fundamental pillar in the development of open innovation processes that enables the generation of new opportunities, new trends, and new products and services (Mention and Asikainen, 2012).

Furthermore, in our data, we found a negative topic of management of open innovation and business models (keyness 174.25 and *p*-value 0.010), which suggests that management plays a decisive role in the success of open innovation, and that traditional or inflexible business systems are major barriers to the development of open innovation (Chesbrough, 2004). Therefore, the support of top management of companies is essential for open innovation. Accordingly, appropriate management and managerial support are a fundamental part of the success of open innovation. However, there are disagreements about management support in traditional business models concerning open innovation (Wåge and Crawford, 2020).

To summarize the discussion and our main findings, Table 9 presents the limits of open innovation according to the analyzed UGC in Twitter, the topics identified in the research, their associated feelings, and previous studies that discussed the results proposed as limits to open innovation. While, when feelings are negative, the limits of open innovation should be taken as factors that companies should avoid, when feelings are positive and neutral, they should be considered good practices to be followed by practitioners.

5.1. *Future research agenda*

Following Jugend et al. (2020) and Ribeiro-Navarrete et al. (2021), in order to encourage further research on open innovation, in Table 10, we outline the questions to be addressed in future studies. The list of questions was compiled based on both our literature review and the results reported in Section 4, as well as based on the topics identified in the present study.

6. Conclusions

In the present study, we applied sentiment analysis, topic modeling, and textual analysis techniques to identify insights and create knowledge about open innovation from a UGC sample extracted from Twitter. Based on the results of our data analysis, we identified a total of eight topics. Of these topics, two topics were negative (Culture and Business Models/Management), three were positive (Communities, Creative projects and Ideas), and three were neutral (Entrepreneurship, Teams, and Technology). Based on the characteristics of each of the topics, we extracted insights that can help to create knowledge on open innovation

Table 9
Limits open innovation according to UGC.

R	Topics	Sentiment	Limits of open innovation	Authors
1	Culture	Negative	- Obsolete and inflexible company cultures that do not provide sufficient commitment to support open innovation as a relevant part of their organization - Wrong internal company corporate culture and strategic transformation models that do not supports open innovation are key factors	Huizingh et al. (2011) Zenobia et al. (2009)
2	Communities	Positive	- The construction of collaborative communities boost open innovation and new ideas in flexible companies - Collaborative communities keeps people motivated and drives the development of ideas focused on open innovation.	Briel and Recker (2017) Martínez-Torres (2014) Fichter (2009)
3	Creative projects	Positive	- Creative projects internal programs are the fundamental axis for open innovation development in companies - Creative projects boost idea sharing, collaborative construction of new ideas, creative thinking, and effective identification of risks. - creative projects should be supported by company management to drive the development of open innovation.	Du et al. (2014) Füller et al. (2012) Boscherini et al. (2010)
4	Entrepreneurship	Neutral	- Entrepreneurial projects and open innovation should be boosted using digital platforms and technologies - Open tools and collaborative activities are drivers for open innovation in entrepreneurship	Schweitzer et al. (2021) Del Vecchio et al. (2018)
5	Ideas	Positive	- Companies supports in terms of action and funding are key elements for open innovation projects success. - Employees positively perceive companies' support and implementation of their ideas concerning open innovation. - The internal management of ideas in companies is essential for employees to feel motivated.	Hofstetter et al. (2021) Igartua et al. (2010) Bughin et al. (2008)
6	Teams	Neutral	- Multidisciplinary and participative teams must be developed to obtain effective open innovation processes - To reach open innovation ideas, teams should be composed of experts from different departments of the company	Markovic et al. (2020) Chatenier et al. (2010)
7	Technology	Neutral	- The adoption of new technological software and hardware allow companies to develop open innovation - Technology is a fundamental pillar that enables the generation of new opportunities, new trends, and new products and services	Bigliardi et al. (2021) Fortunato et al. (2017) Mention and Asikainen (2012)
8	Business Models & Management	Negative	- Management plays a decisive role in the success of open innovation - Traditional or inflexible business systems are major barriers to the development of open innovation. - Appropriate management and managerial support are a fundamental part of the success of open innovation.	Wåge and Crawford (2020) Chesbrough (2004)

Source: the authors

and identify analytical perspectives for the limits and characteristics of open innovation. Based on the results, we also formulated 27 research questions to be addressed in further research on open innovation.

Regarding RQ1 ("*What are the limits of open innovation according to UGC on Twitter?*"), we identified topics that provide a glimpse of the limitations and characteristics according to the database analyzed presented in Table 9 and composed by 20 limitations of open innovation. Of note, these limitations are the basis for the creation of future strategies focused on open innovation communities. Using UGC makes it possible to obtain results that, albeit exploratory, can be used to develop future variables or constructs of quantitative models that seek empirical evidence. In relation to RQ2 (*What is the sentiment of the topics that characterize open innovation according to UGC on Twitter?*), we found that different relevant topics are associated with different feelings analyzed

Table 10
Future research question related to open innovation.

R	Topics	Future Research questions
1	Culture	- What are the characteristics of a resilient and flexible company culture that supports open innovation? - How can obsolete and traditional company cultures be modified so that employees feel that open innovation is important in their company? - How can internal and external values of company cultures be enhanced by the application of new technologies and open innovation projects? - How can external open innovation initiatives promote cultural change and development in a traditional company?
2	Communities	- How can the degree of motivation in open communities driving open innovation to be measured? - What are the acceptance standards of ideas generated in open innovation communities? - How can the perception of belonging to an open community focused on open innovation be measured? - What are the benefits and risks of open innovation-centered project ideation within collaborative communities?
3	Creative projects	- What elements of creativity should be used in creative projects that drive open innovation? - What are the risks associated with managing people in creative projects focused on open innovation? - What are the guidelines for collaborative idea creation in open innovation projects?
4	Entrepreneurship	- Are there common characteristics that link entrepreneurial projects with open innovation? - What is the success factor of entrepreneurs who use open innovation vs. those who do not? - How does the success factor of entrepreneurs affect the ideation of new entrepreneurial projects when open innovation is used?
5	Ideas	- Which actions are most positively perceived by employees when the company supports their open innovation ideas? - What are the appropriate communication channels for ideas focused on open innovation initiatives? - How should new open innovation project ideas be structured and developed? - Is it appropriate to measure open innovation ideas in terms of financials and profitability?
6	Teams	- What should the design of the departments that make up teams prioritize when driving open innovation? - Is it possible to define motivation indicators for the members of a team focused on developing open innovation? - Is it possible to define a framework that regulates the role of external experts in internal open innovation teams?
7	Technology	- Is the use of new technologies and their application related to improvements in open innovation projects? - Should technology be a key factor in the development of open innovation-focused projects? - What are the limitations and risks of using artificial intelligence for the ideation of open innovation projects?
8	Business Models & Management	- How do new generations of managers influence senior management in companies that do not support open innovation? - How can open innovation be included in inflexible business models? - What are the risks of inefficient management in new open innovation projects?

Source: the authors

and discussed in previous sections.

According to our results, Business Culture constitutes one of the main limits of open innovation, as many companies do not support flexible models of business organization. At the same time, collaborative communities in the creation of collective knowledge make a positive contribution to open innovation. Similarly, creative projects driven by idea building with a focus on open innovation were identified as relevant issues for this type of research. Furthermore, our results also highlighted the influence of entrepreneurship as a precursor for the development of open innovation.

Furthermore, as revealed by our results, business support for ideas, funding, and activities to keep idea-generating team members motivated are fundamental to open innovation. At the same time, our results also suggest that teams driving open innovation must be guided by protocols that encourage the acceptance of new technologies, motivation of their proposals, as well as flexibility and adaptation of business models to new products or services implemented within the framework of open innovation initiatives.

Another important insight derived from our data analysis is that technology plays a fundamental role in open innovation processes and is, in most cases, the driving force behind open innovation. Finally, what else is the success of open innovation is central business models and management.

6.1. Theoretical implications

Theoretical implications of the results of the present study are twofold. First, considering the lack of previous research that would investigate open innovation using machine learning methodologies, on the theoretical level, future studies can use the methodology proposed in the present study as a basis for new proposals for knowledge creation and extraction of insights from major social networks.

Second, the themes identified in the present study can be operationalized and explored as variables using quantitative models that seek statistical significance. Said differently, Twitter-based insights and knowledge about open innovation obtained in the present study can become variables for future empirical research. Although the present study is exploratory in nature, the present investigation can serve as a precursor of future quantitative investigations on open innovation, since our aim was not to test hypotheses, but rather to identify variables that can be used in the formulation of future hypotheses in empirical studies.

In this way, the topics identified in the present study can be used as reference in future research that would use construction of statistical models and justify theoretical variables related to open innovation with Partial Least Squares Structural Equation Modeling (PLS-SEM), AMOS (Analysis of Moment Structures), or LISREL (Linear Structural Relations), among others. Using these variables, other researchers can consider the insights identified in the present study to develop research questions and objectives, justify hypotheses, or develop questionnaires.

6.2. Practical implications

The results of the present study are eminently practical. Accordingly, managers of companies or organizations, both public and private, can meaningfully use our results as a guide for the elaboration of new communication or organizational protocols that would foster the development of open innovation in their respective companies. Furthermore, the different feelings identified in this research provide a deeper understanding of employees or external specialists' feelings about open innovation and related topics.

Our results also provide meaningful insights concerning how companies should organize or promote their culture, communities, project management and development, support for entrepreneurial projects, organization and promotion of ideas, structuring and organization of work teams, the role of technology in organizations, as well as the importance of business models and management for open innovation.

Specifically, based on our results and the 8 identified topics, the practitioners can understand the concept of open innovation from different analysis perspectives and not only those related to limitations, risks, and characteristics. More specifically, companies can take these results to improve innovative ideas, promote access to knowledge for the company's stakeholders, encourage innovation among employees, and thus promote the their internal talent, or the development of internal open innovation plans in the organization.

Finally, the CEOs and executives of organizations can propose improvements on their business models and traditional processes by suggesting new ways to boost creativity and generate value through open innovation. Companies can also use open innovation by addressing the research questions proposed in each of the topics identified as relevant for open innovation strategies.

6.3. Limitations

The present study has several limitations. First, we used a non-probability sampling design, so while the results can be taken into account to explain the UGC pieces that are part of the sample, they cannot be generalized to other contexts or samples on the same subject. Second, since all models that work with machine learning are focused on the success of the results in terms of accuracy (so that the more trained a machine learning model is, the better are the results), a limitation of the present study is the number of tweets we used to train our model. Furthermore, another limitation of the present study is that we focused on the data extracted from only one social network (Twitter). Therefore, in further research, it would be necessary to validate our findings on the data from other major social network platforms. Thirdly, the present study is an exploratory investigation that yielded qualitative results. Accordingly, further quantitative research that would quantitatively test the variables identified in the present study is warranted.

Acknowledgments

In gratitude to the Ministry of Science, Innovation and Universities and the European Regional Development. Fund: RTI2018-096295-B-C22.

References

Abouzeedan, A., Hedner, T., 2012. Organization Structure Theories and Open Innovation Paradigm. World Journal of Science, Technology and Sustainable Development.
Adamides, E., Karacapilidis, N., 2020. Information technology for supporting the development and maintenance of open innovation capabilities. J. Innov. & Knowl. 5 (1), 29–38.
Al Amrani, Y., Lazaar, M., El Kadiri, K.E., 2018. Random forest and support vector machine based hybrid approach to sentiment analysis. Procedia Comput. Sci. 127, 511–520.
Alassaf, D., Dabic, M., Daim, T., Shifrer, D., 2020. The impact of open-border organization culture and employees' knowledge, attitudes, and rewards with regards to open innovation: an empirical study. J. Knowl. Manag. 24 (9), 2273–2297. https://doi.org/10.1108/JKM-02-2020-0122. Published online ahead of print.
Bagherzadeh, M., Markovic, S., Bogers, M., 2019. Managing Open Innovation: A Project-Level Perspective. IEEE Transactions on Engineering Management. https://doi.org/10.1109/TEM.2019.2949714. Published online ahead of print:
Bahtar, A.Z., Muda, M., 2016. The impact of User–Generated Content (UGC) on product reviews towards online purchasing–A conceptual framework. Procedia Econ. Finan. 37, 337–342. https://doi.org/10.1016/S2212-5671(16)30134-4.
Barham, H., Dabic, M., Daim, T., Shifrer, D., 2020a. The Role of Management Support for the Implementation of Open Innovation Practices in Firms. Technology in Society, p. 101282. https://doi.org/10.1016/j.techsoc.2020.101282.
Barham, H., Dabic, M., Daim, T., Shifrer, D., 2020b. The role of management support for the implementation of open innovation practices in firms. Technol. Soc. 63, 101282.
Belk, R., 2014. Digital consumption and the extended self. J. Market. Manag. 30 (11–12), 1101–1118. https://doi.org/10.1080/0267257X.2014.939217.
Bhavsar, H., Manglani, R., 2019. Sentiment analysis of Twitter data using Python. Int. Res. J. Eng. Technol. 6 (3), 510–527.
Bigliardi, B., Ferraro, G., Filippelli, S., Galati, F., 2021. The Past, Present and Future of Open Innovation. European Journal of Innovation Management.
Bogers, M., Zobel, A.-K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., Frederiksen, L., Gawer, A., Gruber, M., Haefliger, S., Hagedoorn, J., Hilgers, D., Laursen, K., Magnusson, M.G., Majchrzak, A., McCarthy, I.P., Moeslein, K.M., Nambisan, S., Piller, F.T., Radziwon, A., Rossi-Lamastra, C., Sims, J., Ter Wal, A.L.J., 2017. The open innovation research landscape: established perspectives and emerging themes across different levels of analysis. Ind. Innovat. 24 (1), 8–40.
Bogers, M., Chesbrough, H., Moedas, C., 2018. Open innovation: research, practices, and policies. Calif. Manag. Rev. 60 (2), 5–16.
Boscherini, L., Chiaroni, D., Chiesa, V., Frattini, F., 2010. How to use pilot projects to implement open innovation. Int. J. Innovat. Manag. 14, 1065–1097, 06.
Brunswicker, S., Chesbrough, H., 2018. The Adoption of Open Innovation in Large Firms: practices, Measures, and Risks A survey of large firms examines how firms approach open innovation strategically and manage knowledge flows at the project level. Res. Technol. Manag. 61 (1), 35–45.
Bughin, J., Chui, M., Johnson, B., 2008. The next step in open innovation. McKinsey Q. 4 (6), 1–8.
Cavallo, A., Burgers, H., Ghezzi, A., Van de Vrande, V., 2021. The Evolving Nature of Open Innovation Governance: A Study of a Digital Platform Development in Collaboration with a Big Science Centre. Technovation, p. 102370.
Chatenier, E.D., Verstegen, J.A., Biemans, H.J., Mulder, M., Omta, O.S.F., 2010. Identification of competencies for professionals in open innovation teams. R&d Manag. 40 (3), 271–280.
Chesbrough, H., 2004. Managing open innovation. Res. Technol. Manag. 47 (1), 23–26.
Chesbrough, H., Bogers, M., 2014. Explicating open innovation: clarifying an emerging paradigm for understanding innovation. In: Chesbrough, H., Vanhaverbeke, W., West, J. (Eds.), *New Frontiers In Open Innovation*: 3-28. Oxford University Press, Oxford.
Chesbrough, H., Crowther, A.K., 2006. Beyond high tech: early adopters of open innovation in other industries. R&d Manag. 36 (3), 229–236. https://doi.org/10.1111/j.1467-9310.2006.00428.x.
Chesbrough, H., Lettl, C., Ritter, T., 2018. Value creation and value capture in open innovation. J. Prod. Innovat. Manag. 35 (6), 930–938.
Ciesielska, M., 2018. Nokia on the slope: the failure of a hybrid open/closed source model. Int. J. Enterpren. Innovat. 19 (3), 218–225.
Dabić, M., Lažnjak, J., Smallbone, D., Švarc, J., 2019. Intellectual capital, organisational climate, innovation culture, and SME performance: evidence from Croatia. J. Small Bus. Enterprise Dev. 26 (4), 522–544. https://doi.org/10.1108/JSBED-04-2018-0117.
Dahlander, L., Gann, D.M., 2010. How open is innovation? Res. Pol. 39 (6), 699–709.
Daim, T.U., Chiavetta, D., Porter, A.L., Saritas, O. (Eds.), 2016. Anticipating Future Innovation Pathways through Large Data Analysis. Springer International Publishing, Berlin.
Daradkeh, M.K., 2021. Exploring the usefulness of user-generated content for business intelligence in innovation: empirical evidence from an online open innovation community. Int. J. Enterprise Inf. Syst. 17 (2), 44–70.
Del Vecchio, P., Di Minin, A., Petruzzelli, A.M., Panniello, U., Pirri, S., 2018. Big data for open innovation in SMEs and large corporations: trends, opportunities, and challenges. Creativ. Innovat. Manag. 27 (1), 6–22. https://doi.org/10.1111/caim.12224.
Dhaoui, C., Webster, C.M., Tan, L.P., 2017. Social media sentiment analysis: lexicon versus machine learning. J. Consum. Market. 34 (6), 480–488. https://doi.org/10.1108/JCM-03-2017-2141.
Dhar, V., Chang, E.A., 2009. Does chatter matter? The impact of user-generated content on music sales. J. Interact. Market. 23 (4), 300–307.
Di Domenico, G., Sit, J., Ishizaka, A., Nunan, D., 2021. Fake news, social media and marketing: a systematic review. J. Bus. Res. 124, 329–341. https://doi.org/10.1016/j.jbusres.2020.11.037.
Du, J., Leten, B., Vanhaverbeke, W., 2014. Managing open innovation projects with science-based and market-based partners. Res. Pol. 43 (5), 828–840.
Elia, G., Petruzzelli, A.M., Urbinati, A., 2020. Implementing open innovation through virtual brand communities: a case study analysis in the semiconductor industry. Technol. Forecast. Soc. Change 155, 119994. https://doi.org/10.1016/j.techfore.2020.119994.
Enkel, E., Gassmann, O., Chesbrough, H., 2009. Open R&D and open innovation: exploring the phenomenon. R D Manag. 39 (4), 311–316.
Fichter, K., 2009. Innovation communities: the role of networks of promotors in Open Innovation. R&d Manag. 39 (4), 357–371.
Fischer, D., Prasuhn, J., Strese, S., Brettel, M., 2021. The role of social media for radical innovation in the new digital age. Int. J. Innovat. Manag. 2150075.
Fortunato, A., Gorgoglione, M., Messeni Petruzzelli, A., Panniello, U., 2017. Leveraging big data for sustaining open innovation: the case of social TV. Inf. Syst. Manag. 34 (3), 238–249. https://doi.org/10.1080/10580530.2017.1330000.
Füller, J., Matzler, K., Hutter, K., Hautz, J., 2012. Consumers' creative talent: which characteristics qualify consumers for open innovation projects? An exploration of asymmetrical effects. Creativ. Innovat. Manag. 21 (3), 247–262.
Gabrielatos, C., 2017, October. Clusters of keyness: a principled approach to selecting key items. In: Corpus Linguistics in the South.
García-Quevedo, J., Segarra-Blasco, A., Teruel, M., 2018. Financial constraints and the failure of innovation projects. Technol. Forecast. Soc. Change 127, 127–140.
Gatzweiler, A., Blazevic, V., Piller, F.T., 2017. Dark side or bright light: destructive and constructive deviant content in consumer ideation contests. J. Prod. Innovat. Manag. 34 (6), 772–789.
Greco, M., Grimaldi, M., Cricelli, L., 2018. Benefits and Costs of Open Innovation: the BeCO Framework. Technology Analysis & Strategic Management. https://doi.org/10.1080/09537325.2018.1484442. June, 1–14.
Grover, P., Kar, A.K., Janssen, M., Ilavarasan, P.V., 2019. Perceived usefulness, ease of use and user acceptance of blockchain technology for digital transactions–insights from user-generated content on Twitter. Enterprise Inf. Syst. 13 (6), 771–800.

Guizzo, D., 2019. Discursive strategies in the Keynes-Hayek debate: building a liberal critique. Contrib. Polit. Econ. 38 (1), 12–30. https://doi.org/10.1093/cpe/bzz014.
Hasan, A., Moin, S., Karim, A., Shamshirband, S., 2018. Machine learning-based sentiment analysis for twitter accounts. Math. Comput. Appl. 23 (1), 11.
Hervas-Oliver, J.L., Sempere-Ripoll, F., Boronat-Moll, C., 2021. Technological innovation typologies and open innovation in SMEs: beyond internal and external sources of knowledge. Technol. Forecast. Soc. Change 162, 120338.
Hill, S., Benton, A., Panniello, U., 2019. Talkographics: measuring TV and brand audience demographics and interests from user-generated content. Int. J. Electron. Commer. 23 (3), 364–404. https://doi.org/10.1080/10864415.2019.1619908.
Hiremath, B.N., Patil, M.M., 2020. Enhancing optimized personalized therapy in clinical decision support system using natural language processing. J. King Saud Univ.-Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2020.03.006.
Hirschberg, J., Manning, C.D., 2015. Advances in natural language processing. Science 349 (6245), 261–266. https://doi.org/10.1126/science.aaa8685, 2015.
Ho-Dac, N.N., 2020. The value of online user generated content in product development. J. Bus. Res. 112, 136–146.
Hofstetter, R., Dahl, D.W., Aryobsei, S., Herrmann, A., 2021. Constraining ideas: how seeing ideas of others harms creativity in open innovation. J. Market. Res. 58 (1), 95–114.
Huizingh, E., Conn, S., Torkkeli, M., 2011. Open innovation: state of the art and future perspectives. ISPIM special issue on open innovation. Technovation 31 (1), 2–9.
Igartua, J.I., Garrigós, J.A., Hervas-Oliver, J.L., 2010. How innovation management techniques support an open innovation strategy. Res. Technol. Manag. 53 (3), 41–52.
Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., Zhao, L., 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. Multimed. Tool. Appl. 78 (11), 15169–15211.
Jiang, L., Wang, S., Li, C., Zhang, L., 2016. Structure extended multinomial naive Bayes. Inf. Sci. 329, 346–356.
Jugend, D., Fiorini, P.D.C., Armellini, F., Ferrari, A.G., 2020. Public support for innovation: a systematic review of the literature and implications for open innovation. Technol. Forecast. Soc. Change 156, 119985.
Kim, H., Jang, S.M., Kim, S.H., Wan, A., 2018. Evaluating sampling methods for content analysis of Twitter data. Social Media+ Soc. 4 (2), 2056305118772836.
Klein, A.Z., Magge, A., O'Connor, K., Amaro, J.I.F., Weissenbacher, D., Hernandez, G.G., 2021. Toward using twitter for tracking covid-19: a natural language processing pipeline and exploratory data set. J. Med. Internet Res. 23 (1), e25314.
Krippendorff, K., 2018. Content Analysis: an Introduction to its Methodology. Sage publications.
Krumm, J., Davies, N., Narayanaswami, C., 2008. User-generated content. IEEE Pervasive Comput. 7 (4), 10–11.
Kumar, K.S., Desai, J., Majumdar, J., 2016, December. Opinion mining and sentiment analysis on online customer review. In: 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, pp. 1–4.
Lau, K.W., Wu, Q.H., 2003. Online training of support vector classifier. Pattern Recogn. 36 (8), 1913–1920.
Lehdonvirta, V., Oksanen, A., Räsänen, P., Blank, G., 2021. Social media, web, and panel surveys: using non-probability samples in social and policy research. Pol. Internet 13 (1), 134–155. https://doi.org/10.1002/poi3.238.
Leon, N., 2009. The future of computer-aided innovation. Comput. Ind. 60 (8), 539–550. https://doi.org/10.1016/j.compind.2009.05.010.
Leonardi, P.M., 2014. Social media, knowledge sharing, and innovation: toward a theory of communication visibility. Inf. Syst. Res. 25 (4), 796–816.
Leung, L., 2009. User-generated content on the internet: an examination of gratifications, civic engagement and psychological empowerment. New Media Soc. 11 (8), 1327–1347.
Lexhagen, M., Kuttainen, C., Fuchs, M., Höpken, W., Iaf, H.R.W., 2012. Destination Talk in Social Media: a Content Analysis for Innovation. Advances in Hospitality and Tourism Marketing and Management. Corfu.
Li, G., Liu, F., 2014. Sentiment analysis based on clustering: a framework in improving accuracy and recognizing neutral opinions. Appl. Intell. 40 (3), 441–452.
Lifshitz-Assaf, H., Tushman, M.L., Lakhani, K.R., 2018. A study of NASA scientists shows how to overcome barriers to open innovation. Harv. Bus. Rev. https://hbr.org/2018/05/why-do-open-innovation-efforts-fail-scientists-want-to-solve-problems-themselves March 1st 2020.
Linvill, D.L., McGee, S.E., Hicks, L.K., 2012. Colleges' and universities' use of Twitter: a content analysis. Publ. Relat. Rev. 38 (4), 636–638.
Liu, H., Liu, W., Yoganathan, V., Osburg, V.S., 2021. COVID-19 information overload and generation Z's social media discontinuance intention during the pandemic lockdown. Technol. Forecast. Soc. Change 166, 120600.
Loughran, T., McDonald, B., 2016. Textual analysis in accounting and finance: a survey. J. Account. Res. 54 (4), 1187–1230.
Luca, M., 2015. User-generated content and social media. In: Handbook of Media Economics, vol. 1, pp. 563–592.
Luqman, A., Talwar, S., Masood, A., Dhir, A., 2021. Does enterprise social media use promote employee creativity and well-being? J. Bus. Res. 131, 40–54.
Markovic, S., Bagherzadeh, M., Dubiel, A., Cheng, J., Vanhaverbeke, W., 2020. Do not miss the boat to outside-in open innovation: enable your employees. Ind. Market. Manag. 91, 152–161. https://doi.org/10.1016/j.indmarman.2020.09.003.
Marshall, A., Dencik, J., Singh, R.R., 2021. Open Innovation: Digital Technology Creates New Opportunities. Strategy & Leadership.
Martínez-Torres, M.R., 2014. Analysis of open innovation communities from the perspective of social network analysis. Technol. Anal. Strat. Manag. 26 (4), 435–451.
McEnery, T., Hardie, A., 2013. The History of Corpus Linguistics. The Oxford handbook of the history of linguistics, pp. 727–745.
Mehmet, M.I., Clarke, R.J., 2016. B2B social media semantics: analysing multimodal online meanings in marketing conversations. Ind. Market. Manag. 54, 92–106. https://doi.org/10.1016/j.indmarman.201512.006.
Mention, A.L., 2011. Co-operation and co-opetition as open innovation practices in the service sector: which influence on innovation novelty? Technovation 31 (1), 44–53.
Mention, A.L., Asikainen, A.L., 2012. Innovation & productivity: investigating effects of openness in services. Int. J. Innovat. Manag. 16, 1240004, 03.
Minowa, Y., Belk, R.W., 2017. Ad hoc Japonisme: how national identity rhetorics work in Japanese advertising. Consum. Mark. Cult. 20 (4), 329–349. https://doi.org/10.1080/10253866.2016.1239085.
Moretti, F., Biancardi, D., 2020. Inbound open innovation and firm performance. J. Innov. & Knowl. 5 (1), 1–19.
Muninger, M.I., Hammedi, W., Mahr, D., 2019. The value of social media for innovation: a capability perspective. J. Bus. Res. 95, 116–127.
Nguyen, B., Yu, X., Melewar, T.C., Chen, J., 2015. Brand innovation and social media: knowledge acquisition from social media, market orientation, and the moderating role of social media strategic capability. Ind. Market. Manag. 51, 11–25.
Noh, H., Lee, S., 2020. What constitutes a promising technology in the era of open innovation? An investigation of patent potential from multiple perspectives. Technol. Forecast. Soc. Change 157, 120046.
North-Kim, A.J., Johnson, K.K., 2016. Power of consumers using social media: examining the influences of brand-related user-generated content on Facebook. Comput. Hum. Behav. 58, 98–108.
Östman, J., 2012. Information, expression, participation: how involvement in user-generated content relates to democratic engagement among young people. New Media Soc. 14 (6), 1004.
Ozcan, S., Suloglu, M., Sakar, C.O., Chatufale, S., 2021. Social media mining for ideation: identification of sustainable solutions and opinions. Technovation 107, 102322.
Palguna, D.S., Joshi, V., Chakaravarthy, V., Kothari, R., Subramaniam, L.V., 2015. June). Analysis of sampling algorithms for twitter. In: Twenty-Fourth International Joint Conference on Artificial Intelligence.
Park, I., Yoon, B., Kim, S., Seol, H., 2019. Technological Opportunities Discovery for Safety through Topic Modeling and Opinion Mining in the Fourth Industrial Revolution: the Case of Artificial Intelligence. IEEE Transactions on Engineering Management.
Pelka, B., Kaletka, C., 2011. Web 2.0 revisited: user-generated content as a social innovation. Int. J. Innovat. Sustain. Dev. 5 (2–3), 264–275.
Pilav-Velic, A., Jahic, H., 2021. The adoption of inbound open innovation practices in developing countries: empirical evidence from the manufacturing sector. Eur. J. Innovat. Manag. https://doi.org/10.1108/EJIM-11-2020-0460.
Pojanapunya, P., Todd, R.W., 2018. Log-likelihood and odds ratio: keyness statistics for different purposes of keyword analysis. Corpus Linguist. Linguistic Theory 14 (1), 133–167.
Rayson, P., Garside, R., 2000. October). Comparing corpora using frequency profiling. In: The Workshop on Comparing Corpora, pp. 1–6.
Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V.A., Boyd-Graber, J., 2015. Beyond LDA: exploring supervised topic modeling for depression-related language in Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: from Linguistic Signal to Clinical Reality, pp. 99–107.
Ribeiro-Navarrete, S., Saura, J.R., Palacios-Marqués, D., 2021. Towards a new era of mass data collection: assessing pandemic surveillance technologies to preserve user privacy. Technol. Forecast. Soc. Change 167, 120681. https://doir.org/10.1016/j.techfore.2021.120681.
Roberts, D.L., Piller, F.T., 2016. Finding the right role for social media in innovation. MIT Sloan Manag. Rev. 57 (3), 41–47.
Roma, P., Aloini, D., 2019. How does brand-related user-generated content differ across social media? Evidence reloaded. J. Bus. Res. 96, 322–339.
Saura, J.R., Palacios-Marqués, D., Iturricha-Fernández, A., 2021. Ethical Design in Social Media: assessing the main performance measurements of user online behavior modification. J. Bus. Res. 129 (May 2021), 271–281. https://doi.org/10.1016/j.jbusres.2021.03.001.
Saura, J.R., Ribeiro-Soriano, D., Palacios-Marqués, D., 2021a. Using data mining techniques to explore security issues in smart living environments in Twitter. Comput. Commun. https://doi.org/10.1016/j.comcom.2021.08.021.
Saura, J.R., Ribeiro-Soriano, D., Palacios-Marqués, D., 2021b. Setting B2B Digital Marketing in Artificial Intelligence-based CRMs: a review and directions for future research. Ind. Market. Manag. 98, 161–178. https://doi.org/10.1016/j.indmarman.2021.08.006JR.
Schmunk, S., Höpken, W., Fuchs, M., Lexhagen, M., 2013. Sentiment analysis: extracting decision-relevant knowledge from UGC. In: *Information And Communication Technologies in Tourism* 2014. Springer, Cham, pp. 253–265.
Schweitzer, F., Palmié, M., Gassmann, O., Kahlert, J., Roeth, T., 2021. Open Innovation for Institutional Entrepreneurship: How Incumbents Induce Institutional Change to Advance Autonomous Driving. R&D Management.
Shan, W., Qiao, T., Zhang, M., 2020. Getting more resources for better performance: the effect of user-owned resources on the value of user-generated content. Technol. Forecast. Soc. Change 161, 120318.
Short, J.C., Broberg, J.C., Cogliser, C.C., Brigham, K.H., 2010. Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. Organ. Res. Methods 13 (2), 320–347. https://doi.org/10.1177/1094428109335949.
Sibona, C., Walczak, S., White Baker, E., 2020. A guide for purposive sampling on twitter. Commun. Assoc. Inf. Syst. 46 (1), 22.

Sihombing, S.O., 2017. Predicting intention to share news through social media: an empirical analysis in Indonesian youth context. Bus. Econ. Horiz. 13 (4), 468–477. https://doi.org/10.15208/beh.2017.32.

Tajbakhsh, M.S., Bagherzadeh, J., 2019. Semantic knowledge LDA with topic vector for recommending hashtags: twitter use case. Intell. Data Anal. 23 (3), 609–622.

Tandon, C., Revankar, S., Parihar, S.S., 2021. How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics. Int. J. Inform. Manag. Data Insights 1 (2), 100035.

Täuscher, K., Bouncken, R., Pesch, R., 2020. Gaining legitimacy by being different: optimal distinctiveness in crowdfunding platforms. Acad. Manag. J. https://doi.org/10.5465/amj.2018.0620.

Utama, H., 2019. November). Sentiment analysis in airline tweets using mutual information for feature selection. In: 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). IEEE, pp. 295–300.

Van Dijck, J., 2009. Users like you? Theorizing agency in user-generated content. Media, culture & society 31 (1), 41–58.

Vanhaverbeke, W., Roijakkers, N., Lorenz, A., Chesbrough, H., 2017. The importance of connecting open innovation to strategy. In: Strategy and Communication for Innovation. Springer, pp. 3–15.

von Briel, F., Recker, J.C., 2017. Lessons from a failed implementation of an online open innovation community in an innovative organization. MIS Q. Exec. 16 (1), 35–46.

Wåge, D., Crawford, G.E., 2020. Innovation in digital business models. In: Managing Digital Open Innovation, pp. 35–45.

West, J., Bogers, M., 2014. Leveraging external sources of innovation: a review of research on open innovation. J. Prod. Innovat. Manag. 31 (4), 814–831.

West, J., Salter, A., Vanhaverbeke, W., Chesbrough, H., 2014. Open innovation: the next decade. Res. Pol. 43, 805–811.

Widaretna, T., Tirtawangsa, J., Romadhony, A., 2021, August. Hoax identification on tweets in Indonesia using Doc2Vec. In: 2021 9th International Conference on Information and Communication Technology (ICoICT). IEEE, pp. 456–461.

Wilson, R.T., Baack, D.W., Till, B.D., 2015. Creativity, attention and the memory for brands: an outdoor advertising field study. Int. J. Advert. 34 (2), 232.

Xuanyang, X., Yuchang, G., Shouhong, W., Xi, L., 2006. Computer aided detection of SARS based on radiographs data mining. In: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. IEEE, pp. 7459–7462. https://doi.org/10.1109/IEMBS.2005.1616237.

Yang, M., Chen, K., Miao, Z., Yang, X., 2014, December. Cost-effective user monitoring for popularity prediction of online user-generated content. In: 2014 IEEE International Conference on Data Mining Workshop. IEEE, pp. 944–951.

Zafar, M.B., Bhattacharya, P., Ganguly, N., Gummadi, K.P., Ghosh, S., 2015. Sampling content from online social networks: comparing random vs. expert sampling of the twitter stream. ACM Trans. Web 9 (3), 1–33.

Zenobia, B., Weber, C., Daim, T., 2009. Artificial markets: a review and assessment of a new venue for innovation research. Technovation 29 (5), 338–350.

Zhang, W., Kang, L., Jiang, Q., Pei, L., 2018. From buzz to bucks: the impact of social media opinions on the locus of innovation. Electron. Commer. Res. Appl. 30, 125–137.

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., Li, X., 2011. April). Comparing twitter and traditional media using topic models. In: European Conference on Information Retrieval. Springer, Berlin, Heidelberg, pp. 338–349.

Zhu, X., Xiao, Z., Dong, M.C., Gu, J., 2019. The fit between firms' open innovation and business model for new product development speed: a contingent perspective. Technovation 86, 75–85. https://doi.org/10.1016/j.technovation.2019.05.005.