frontiers in Sociology

TYPE Original Research
PUBLISHED 09 September 2026
DOI 10.3389/fsoc.2026.1874323

Check for updates

OPEN ACCESS

EDITED BY
Fandi Achmad,
Telkom University, Indonesia

REVIEWED BY
Philipp Altmann,
Central University of Ecuador, Ecuador
Muhammad Rifqi Maarif,
Universitas Tidar, Indonesia

*CORRESPONDENCE
Lenin Quiñones Huatangari
lenin.quinones@untrm.edu.pe

RECEIVED 06 May 2026
REVISED 31 July 2026
ACCEPTED 24 August 2026
PUBLISHED 09 September 2026

CITATION
Quiñones Huatangari L, Díaz Cabrera R, Alvarado Carrasco N, Fernández Rosillo F, Maicelo Guevara JL, Cruz Caro O and Salazar Serván S (2026) Analysis of the minutes of rondas campesinas using text mining techniques.
Front. Sociol. 11:1874323.
doi: 10.3389/fsoc.2026.1874323

COPYRIGHT
© 2026 Quiñones Huatangari, Díaz Cabrera, Alvarado Carrasco, Fernández Rosillo, Maicelo Guevara, Cruz Caro and Salazar Serván. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Analysis of the minutes of rondas campesinas using text mining techniques

Lenin Quiñones Huatangari[1]*, Raúl Díaz Cabrera[2], Nicanor Alvarado Carrasco[3], Frank Fernández Rosillo[4], Jorge Luis Maicelo Guevara[5], Omer Cruz Caro[6] and Segundo Salazar Serván[7]

[1]Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Facultad de Ingeniería Zootecnista y Biotecnología, Escuela Profesional de Ingeniería de Ciencia de Datos e Inteligencia Artificial, Chachapoyas, Peru, [2]Instituto de Educación Superior Privado "CERTUS", Carrera de Diseño y Desarrollo de Software, Lima, Peru, [3]Universidad Nacional de Jaén, Instituto de Investigación en Ciencias Sociales, Jaén, Cajamarca, Peru, [4]Universidad Nacional de Jaén, Instituto de Investigación en Ciencia de Datos, Jaén, Cajamarca, Peru, [5]Pontificia Universidad Católica del Perú, Escuela Profesional de Economía, Lima, Peru, [6]Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Oficina de Gestión de la Calidad, Chachapoyas, Peru, [7]Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Dirección de Admisión y Asuntos Académicos, Chachapoyas, Peru

The rondas campesinas are rural community justice organizations that emerged among impoverished indigenous communities in the Andes and are the driving force behind the largest and most influential grassroots movement in northeastern Peru. The aim of the research was to unveil the characteristics, thematic approaches and working dynamics of the rondas campesinas using text mining as a technique for information analysis. Seventy-four minutes from Huaquillas, Namballe and San Jose de Lourdes in the province of San Ignacio, department of Cajamarca were analyzed, covering the period from 2012 to 2024. The analysis included five steps: data collection and pre-processing by eliminating stopwords and tokenization, keyword identification using the word cloud technique, word quantification using the Term Frequency-Inverse Document Frequency method, latent topic analysis through Latent Dirichlet assignment and vector representation through modeling with Word2Vec. These steps allowed the identification of community problems, social control actions and emerging trends in community organization. The text mining technique is a tool to analyze minutes not only of rondas campesinas but also of organizations or institutions that want to know with certainty the thematic approaches they develop.

KEYWORDS
artificial intelligence, governance, intercultural approach, legal pluralism, topical analysis

1 Introduction

Throughout the history of the formation of civilizations around the world, various mechanisms of social dynamics have been observed, circumscribed by the search for the common good as a primary goal (Touraine, 1985), has led to the grouping of individuals around ideas, philosophies and premises of collective interest (Edelman, 2001), that have governed the consensus and concertation of nations, which has contributed to the consolidation of modern societies (Melucci, 1984). Therefore, there are own or external factors that bring people together, territorially or socially, with common rather than individual objectives, interests and rules (Lázaro Aquino, 2024). This social dynamism due to the inherent need for development

and growth to satisfy the lifestyle of citizens causes political, geopolitical and technological phenomena, which in some cases contradict an established order, violating common law and prioritizing supra-individual interests (Hamel and Maheu, 2001); where organized civil society seeks and finds new mechanisms that contribute to being an effective alternative in established struggles (Brisman et al., 2016). This is why social movements develop under different geographical, political and situational circumstances, where the search for equity and justice of the interest group or civil society takes precedence and converges (Giugni, 1999). This constitutes a challenge in the study, understanding and addressing of social conflicts based on historical and current data (Doerr et al., 2013), where new technological trends (Rezaev and Tregubova, 2018) contribute to and play an important role in processing and searching for recurring patterns or behaviors in these events (Olsher, 2015).

In South America, social movements and the development of collectives are the product of political phenomena with a lack of precision and effectiveness in decision-making on issues that concern societies (Inclán, 2018). Rural social movements develop in contexts of historical inequalities, territorial exclusion, and asymmetrical relationships between communities and states (Lopez-Flores, 2024). From this perspective, internal colonialism helps explain why methods of domination over indigenous and rural communities still exist within nations (Casanova, 1965). On the other hand, the coloniality of power explains why social, political, and economic hierarchies inherited from the colonial order persist (Quijano, 2000). Furthermore, the physical and administrative distance between rural communities and government institutions fosters the development of methods of collective collaboration (Davis, 1999). In this context, the "rondas campesinas" emerge as expressions of community self-management related to territorial protection, social regulation, and the building of collective consensus.

On 27 April 1990, peasants from northern Peru stormed the streets of Chota, department of Cajamarca, with the intention of creating other forms of non-state power and justice (Acuña, 2019). The rondas campesinas are rural organizations formed in Cajamarca (Chillihuani Ttito, 2020) of community justice (Nuñez Palomino, 1996), a mass movement of night watchman patrols emerged among impoverished indigenous communities in the Andes (Escalante, 2020); driving the largest and most influential grassroots movement in north-eastern Peru (Shanee, 2019). The minutes of rondas campesinas are important documents that serve as instruments of legitimacy, transparency and collective memory. Therefore, the Peruvian state recognizes them as organizations that can support rural communities (Piccoli, 2009). Plenary Agreement N° 1/2009/CJ-116 is a contribution of the Supreme Court of Peru that contains normative elements to concretize the coordination between the special communal jurisdiction and the ordinary jurisdiction, given the constitutional and conventional mandate that sustains the recognition of the jurisdictional function of the peasant patrols in Peru (Alarcón-Requejo, 2023).

The production, modification and transformation of data are the result of scientific and technical processes, reflecting human values and social relations (Kitchin, 2021). Data have been used as tools and weapons to argue for what is true (Wiggins and Jones, 2023). However, technology has driven their mass creation (Yamin, 2019), facilitating storage (Feldman and Sanger, 2006) and becoming complex (Sawitri, 2024); more than 80% of them are unstructured or semi-structured (Talib et al., 2016) and are in text form (Patel, 2017). It is therefore necessary to extract value and knowledge from these data sets (Elgendy and Elragal, 2014). Advances in the fields of mathematics, computer science, machine learning and natural language processing can collectively capture, classify and interpret words and their contexts (Xiaojun, 2011).

Text mining (TM) is defined as the process of extracting meaningful information and hidden patterns from large volumes of unstructured textual data (del Castillo Zayas and Leiva Mederos, 2007; Markellos, 2009; Markellos, 2009). Its growth has been driven by technological advances (Zanini and Dhawan, 2015), there is software available to employ TM (Percha, 2021), some of the open source tools: KNIME, RapidMiner, Weka, R, and Python (Hofmann and Chisholm, 2016). TM has a range of applications as an information analysis technique (Qamar and Raza, 2024): It reveals the characteristics, thematic focus and working dynamics of a research group (Pérez and Reyes, 2022), understands how citizens perceive climate change and how it impacts on the urban environment and changes the landscape (Choi et al., 2022), recognizes, defines and selects problems (Gyódi et al., 2023), generates a methodology for quantifying socio-political conflicts related to natural resource exploitation through newspaper reports (Palazzo, 2017), assesses perceptions of the environmental, social and governance visions of companies in line with the Sustainable Development Goals (Yoon et al., 2023), analyses news from the Russian news agency RIA Novosti and the Ukrainian news agency during the Russian invasion of Ukraine (Ptaszek et al., 2024), scans data sources for potential threats, trends and accelerates forecasting process (de Boer et al., 2019), extracts health outcomes for the development of prognostic models (Grotenhuis et al., 2024), elucidates the analytical bias of social networking and TM techniques (Singh et al., 2017).

Within this framework, the present study aims to reveal the characteristics, thematic focuses, and organizational dynamics of the rondas campesinas located in northern Peru, through the analysis of their minutes using TM techniques. The study is justified because these minutes are not merely administrative records, but social documents through which communities express agreements, concerns, forms of authority, mechanisms of social control, and claims related to territorial defense, the environment, mining, and relations with local authorities. From the perspective of the sociology of law, analyzing these records makes it possible to understand how rondas campesinas produce and legitimize communal norms, organize collective memory, and sustain their own forms of justice in dialogue or tension with the state legal order. Therefore, the use of tools such as TF-IDF, LDA, and Word2Vec not only provides a methodological strategy for processing unstructured information, but also makes visible discursive patterns that help interpret legal pluralism, territorial governance, and collective action in rural contexts of northern Peru.

2 Materials and methods

2.1 Research framework

The minutes of rondas campesinas are documentary records that function as instruments of legitimacy, transparency, and collective memory. They register community meetings, issues discussed, interventions, agreements, commitments, and decisions related to local organization, social control, conflict management, authorities, mining, environmental concerns, and territorial defense. Although their structure varies according to locality and year, they usually include a

heading, date, place of meeting, identification of the ronda or local base, participants or authorities, agenda, development of the issues discussed, final agreements, and signatures.

In order to analyze the minutes through TM, 74 min were collected from the rondas campesinas of the hamlet of Huaquillas and the districts of Namballe and San José de Lourdes in the province of San Ignacio, department of Cajamarca. The corpus corresponds to the documents provided by the representatives of these organizations and authorized for scanning and analysis, covering the period from 4 March 2012 to 14 March 2024, which provides a broad temporal coverage for an exploratory documentary analysis.

2.2 Methodology

Figure 1 presents the analysis process in five steps adapted from Han et al. (2024): (1) data collection and pre-processing; (2) identification of dominant keywords through word cloud visualization; (3) keyword quantification using Term Frequency-Inverse Document Frequency (TF-IDF); (4) latent topic analysis using Latent Dirichlet Allocation (LDA); and (5) vector representation through Word2Vec modeling and t-SNE visualization.

2.2.1 Step 1: data collection and pre processing

Copies of the minutes were provided by the heads of the peasant patrols of Huaquillas (20), Namballe (18), and San José de Lourdes (36). They were scanned and transcribed with artificial intelligence (ChatGPT Plus) (Fones and Fones, 2025), then reviewed individually to ensure the correct transcription of the information. The collected data were synthesized into CSV files.

Once the data collection was completed, the cleaning process was carried out, which included the elimination of stopwords, punctuation marks, numerical values, etc., from the data (Sakthi Vel, 2021). This procedure divides the text into smaller units called tokens, generally corresponding to individual words. This stage is fundamental, as it allows for a more precise interpretation and analysis of the information content (Manning et al., 2009). Both tokenization and empty word removal are important processes in text analysis that allow researchers and practitioners to focus on the meaningful words and phrases in textual data (Alaminos-Fernández, 2003).

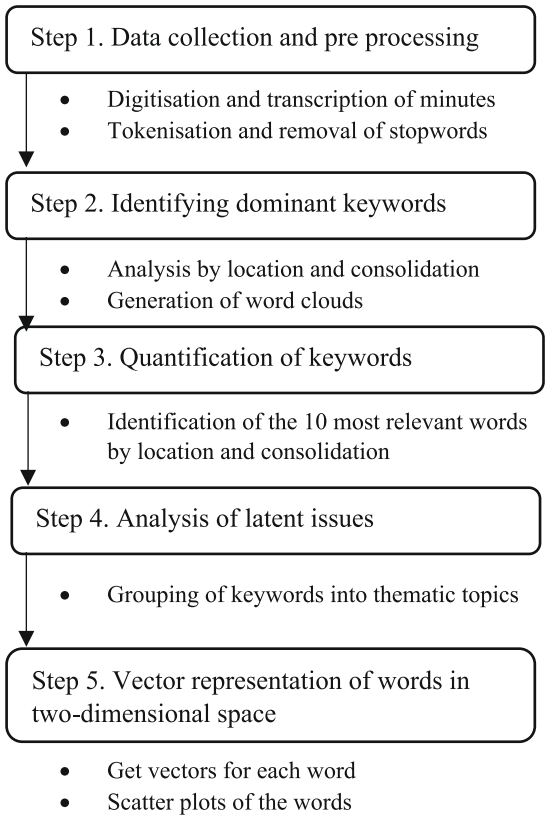


FIGURE 1
Actions in each of the five steps. The methodological workflow was applied to the original Spanish-language minutes. The source corpus was not translated prior to preprocessing or text-mining analysis.

2.2.2 Step 2: identifying dominant keywords

The word cloud technique was applied individually to the databases obtained from the records of Huaquillas, Namballe, and San José de Lourdes; as well as to the dataset integrating the three localities to visualize and summarize volumes of text, highlighting the most frequent and relevant words (Heimerl et al., 2014). For this purpose, the Python 3.11, wordcloud library was used to display the relevant words in each of the contexts analyzed (Dong et al., 2022).

2.2.3 Step 3: quantification of keywords

The Term Frequency-Inverse Document Frequency (TF-IDF) method was used to calculate the weights of the characteristic words within the text corpus (Kim and Gil, 2019). TF-IDF is a weighted measure to assess the importance of a word in a set of documents, taking into account the frequency of the word in the documents and the rarity of the word in the whole set of documents (Salton, 1983). It facilitates the identification of the most relevant words in the location databases, providing a significant advantage for a better interpretation of the results (Yun-tao et al., 2005). The 10 most relevant words from each database belonging to each site, as well as from the global database integrating the three sites, were ordered in descending order.

2.2.4 Step 4: analysis of latent issues

Latent Dirichlet allocation (LDA) is a generative probabilistic model applied to a corpus (Blei et al., 2003). It consists of representing documents as random mixtures of latent topics, where each topic is characterized by a specific distribution of words (Blei et al., 2003). These thematic clusters allow for a more precise identification of emerging trends in each of the sites analyzed (Soliman et al., 2023). Once the sets of words have been defined, a title is assigned to each topic in order to facilitate their interpretation.

For the LDA analysis, independent models were estimated for each documentary subset: Huaquillas, Namballe, San Jose de Lourdes, and the integrated corpus of the three localities. The number of topics was defined separately for each subset, considering differences in corpus size, lexical composition, and thematic density. Candidate models with different values of k were reviewed, and the final number of topics was selected through qualitative validation based on three

criteria: semantic coherence of the most representative words, low redundancy between topics, and interpretability of the resulting clusters in relation to the documentary content of the minutes. Based on this procedure, three topics were retained for Huaquillas, three for Namballe, four for San José de Lourdes, and five for the integrated corpus.

2.2.5 Step 5: vector representation of words in two-dimensional space

The vector representation of words was performed using Word2Vec, a model based on shallow neural networks that generates word vectors from a text corpus (Mikolov et al., 2013). In this study, the embedding process was operationalized at the unigram level. After preprocessing, which included the removal of stopwords, punctuation marks, numerical values, personal names, and non-relevant formal elements, the retained tokens were used to train the Word2Vec model. Bigrams and higher-order n-grams were not included at this stage, since the objective was to examine semantic proximity among individual terms within the minutes.

The model was implemented using the Python gensim library and trained on the documentary corpus from Huaquillas, Namballe, and San José de Lourdes. Parameters such as window size, embedding dimensions, and minimum word frequency were adjusted during model training (Desai et al., 2022). After generating the word embeddings, t-SNE was applied to reduce the vector representations to a two-dimensional space, facilitating visual interpretation (Taskesen and Reinders, 2016). In the resulting maps, each point represents a term retained in the Word2Vec vocabulary after preprocessing, and spatial proximity indicates semantic proximity within the analyzed corpus (Pérez-Serrano et al., 2022). The interpretation focused on terms with greater analytical relevance according to their recurrence, TF-IDF values, and relationship with the LDA topics.

3 Results

Taking into account the objective of the study, which was to reveal the characteristics, thematic approaches and work dynamics of the rondas campesinas in three places in the province of San Ignacio in the Department of Cajamarca in Peru, using TM as a technique for analyzing information, the results are presented based on the methodology employed.

The word cloud procedure made it possible to identify the most recurrent concepts in the documents examined (Jayashankar and Sridaran, 2017). In the dataset covering the three localities (Figure 2a), the words "project," "environment," "development," "representative," and "means" stand out, indicating that the topics addressed are related to development initiatives, environmental concerns, and the participation of community stakeholders. In Huaquillas (Figure 2b), terms related to the management of mining projects, consultation processes, and the protection of natural resources stand out, such as "environment," "development," "project," "means," "consultation," and "model." In Namballe (Figure 2c), the words "mining," "mayor," "project," "agreements," "commitment," and "support" stand out. This reflects an agenda linked to mining activity. It also reflects coordination with local authorities and the establishment of collective commitments. In San José de Lourdes (Figure 2d), a greater emphasis is observed on the terms "representative," "importance," "project," "assembly," "strengthen," "mining," "agreements," and "organization." These terms focus on communal representation, organizational strengthening, and decision-making regarding mining and territorial matters.

Furthermore, keywords were quantified using the TF-IDF technique. As shown in Table 1, the concept "development" obtained the highest weight in Huaquillas, with a value of 0.1409, and in the integrated dataset of the three localities, with 0.0480. In Namballe, the most important word was "mining," with a value of 0.0702, whereas in San José de Lourdes the word "assembly" stood out, with a value of 0.0563. These results have been complemented by the information presented in the word clouds, which show that discursive priorities vary by locality: "development" and "consultation" in Huaquillas, "mining" and "coordination with authorities" in Namballe, and "assembly-based organization" in San José de Lourdes.

The LDA analysis made it possible to identify five key themes in each of the documents from Huaquillas, Namballe, San José de Lourdes, and the three localities combined. The themes were labeled based on the words with the highest probability of occurrence within each cluster and their coherence with the content of the minutes. Table 2 presents the results and reveals the presence of themes related to mining projects, environmental advocacy, consultation processes, coordination with authorities, the establishment of agreements, and the strengthening of communal organization.

The issues identified reveal disparities in the concerns documented by the rondas campesinas across each jurisdiction. In Huaquillas, the predominant themes are consultation processes, mining development models, and the defense of water, land, and other natural resources. In Namballe, the themes are primarily related to mining activity, coordination with local authorities, and mobilization against illegal mining in the rivers. In San José de Lourdes, assemblies, agreements, project management, and the strengthening of communal organization take on greater relevance. In the integrated corpus, these contents converge into themes related to environmental defense, collective participation, engagement with authorities, and consultation on mining development projects.

To examine the semantic relationships among the terms retained after preprocessing, Word2Vec was used together with the t-SNE technique. Figure 3 shows the distribution of the words in the analyzed dataset in two dimensions. In this representation, the proximity between terms indicates that they tend to be used in similar contexts within the minutes, reflecting a relationship of compatibility and frequency of use. However, due to the characteristics of t-SNE, the interpretation focuses primarily on local clusters rather than on the absolute distances between all points in the space. The highlighted terms were selected based on their recurrence, the values obtained through TF-IDF, and their relationship to the themes identified through LDA.

In the dataset covering the three localities (Figure 3a), a proximity can be observed among the words "consultation," "participation," and "mining," reflecting the connection between participatory processes



FIGURE 2
Word clouds of the corpora analyzed by locality: **(a)** Integrated corpus of the three localities; **(b)** Huaquillas; **(c)** Namballe; and **(d)** San José de Lourdes. The word clouds were generated from the original Spanish-language minutes. The terms were not translated prior to analysis in order to preserve their original frequencies and contextual meanings and to avoid potential translation-induced bias.

TABLE 1 TF-IDF of the Huaquillas, Namballe, San José de Lourdes and the three locations.

Huaquillas		Namballe		San Jose de Lourdes		Three locations	
Word	TF-IDF	Word	TF-IDF	Word	TF-IDF	Word	TF-IDF
Desarrollo	0.1409	Mineria	0.0702	Asamblea	0.0563	Desarrollo	0.0480
Consulta	0.1172	Alcalde	0.0623	Congreso	0.0494	Ambiente	0.0435
Modelo Desarrollo	0.1094	Ilegal	0.06	Importancia	0.0456	Medio	0.0419
Modelo	0.1050	Minera	0.0596	Organizacion	0.0453	Medio Ambiente	0.0404
Proyecto Minero	0.0933	Rio	0.0581	Ambiente	0.0444	Proyecto	0.0400
Participacion	0.0753	Apoyo	0.0546	Fortalecer	0.0441	Participacion	0.0373
Participacion Ciudadana	0.0726	Municipalidad	0.0541	Proyectos	0.0440	Consulta	0.0373
Consulta Participacion	0.0720	Conocer	0.0515	Mineria	0.0427	Mineria	0.0366
Ciudadana	0.0701	Acuerdos	0.0504	Participacion	0.0421	Modelo Desarrollo	0.0356
Taller	0.0695	Maquinaria	0.0492	Medio Ambiente	0.0411	Modelo	0.0348
Defensa	0.0650	Proyecto	0.0477	Medio	0.0411	Minero	0.0341

TF-IDF values were calculated from the original Spanish-language corpus. Keywords are therefore presented in Spanish, as prior translation could have altered tokenisation, term frequencies, and the resulting relevance weights.

TABLE 2 Suggested topics from the minutes of the rondas campesinas by LDA.

Key words	Suggested topic
Huaquillas	
0.051 "desarrollo" + 0.042 "consulta" + 0.038 "modelo" + 0.030 "minero" + 0.028 "participación"	Participatory consultation and mining development models
0.025 "agua" + 0.025 "defensa" + 0.025 "tierra" + 0.025 "aire" + 0.020 "minero"	Defense of water, land, and air against mining impacts
0.023 "defensa" + 0.021 "asociación" + 0.017 "acuerdos" + 0.017 "minero" + 0.015 "asamblea"	Collective agreements and commitments of mining-affected groups
0.029 "lucha" + 0.021 "defensa" + 0.021 "apoyo" + 0.016 "defender" + 0.015 "alcalde"	Collective struggle and institutional support for environmental defense
0.023 "minero" + 0.019 "desarrollo" + 0.019 "afectados" + 0.016 "asociación" + 0.016 "tierra"	Mining development and protection of affected territories
Namballe	
0.021 "congreso" + 0.019 "maquinaria" + 0.017 "procedió" + 0.016 "máquina" + 0.016 "explicó"	Congress proceedings and machinery management
0.039 "conocer" + 0.039 "compañeros" + 0.033 "minería" + 0.027 "asistir" + 0.014 "policía"	Community participation and commitments regarding mining
0.022 "alcalde" + 0.020 "informe" + 0.016 "desarrollo" + 0.016 "minera" + 0.015 "acuerdos"	Coordination with local authorities on mining and development
0.052 "proyecto" + 0.033 "apoyo" + 0.033 "lucha" + 0.029 "alcalde" + 0.020 "minería"	Collective support and struggle concerning mining projects
0.042 "minería" + 0.038 "río" + 0.034 "ilegal" + 0.022 "acuerdos" + 0.022 "participación"	Mobilization and agreements against illegal mining in rivers
San José de Lourdes	
0.017 "acuerdos" + 0.015 "informe" + 0.013 "ronderos" + 0.012 "congreso" + 0.012 "minería"	Ronda agreements and reports on mining issues
0.020 "actividades" + 0.018 "importancia" + 0.017 "fortalecer" + 0.016 "locales" + 0.011 "comunal"	Strengthening local activities and communal resources
0.019 "importancia" + 0.017 "asamblea" + 0.016 "proyectos" + 0.015 "fortalecer" + 0.014 "gestión"	Assembly management and organizational strengthening
0.020 "minería" + 0.015 "asamblea" + 0.014 "participación" + 0.012 "ambiente" + 0.010 "representante"	Community participation in mining and environmental decisions
0.022 "congreso" + 0.017 "acuerdos" + 0.012 "organización" + 0.010 "mejorar" + 0.010 "fortalecer"	Congresses and commitments for organizational development
Three locations	
0.017 "proyecto" + 0.015 "alcalde" + 0.013 "minería" + 0.012 "lucha" + 0.011 "apoyo"	Mining projects, local authorities, and collective action
0.027 "ambiente" + 0.026 "medio" + 0.017 "minería" + 0.014 "alcalde" + 0.012 "lucha"	Environmental defense in the face of mining activity
0.013 "explicó" + 0.011 "manera" + 0.009 "tema" + 0.007 "ilegal" + 0.007 "río"	Community discussion of illegal mining and river-related problems
0.012 "asamblea" + 0.012 "acuerdos" + 0.012 "organización" + 0.011 "fortalecer" + 0.009 "participación"	Organizational strengthening and collective participation
0.036 "desarrollo" + 0.026 "modelo" + 0.026 "consulta" + 0.025 "proyecto" + 0.025 "minero"	Participatory consultation on mining development projects

The keywords were retained in Spanish because the LDA models were fitted to the original-language corpus. The suggested topic labels were subsequently expressed in English for reporting purposes. No prior translation of the minutes was performed, thereby preserving the original word distributions and thematic structure.

and the management of mining projects. Furthermore, the words "assembly," "mayor," and "environment" are located in a nearby sector, which reflects the involvement of authorities and communal deliberation spaces in environmental matters. The presence of other terms such as "defense," "mining," "agreements," and "water" in different areas of the semantic space demonstrates that the defense of the territory, the use of natural resources, and the adoption of agreements are recurring themes. However, these do not necessarily cluster into a single category.

In Huaquillas (Figure 3b), the words "participatory," "participation," "citizen," "workshop," and "mining" appear close to one another, suggesting a relationship between community participation, the venues where training is provided, and the manner in which mining activity is conducted. Furthermore, a close relationship is observed between water, defense, and mayor, reflecting the debate over the protection of water resources and the involvement of local authorities. The term "model" appears in another sector of the map, associated with content on development and consultation, which is identified through other analytical techniques.

In Namballe (Figure 3c), a concentration of terms related to mining, the river, and illegality is observed, suggesting the recurrence of issues associated with illegal mining and its potential repercussions for water resources. The terms "mine," "agreements," and "take" appear close together, indicating the adoption of decisions regarding this activity, which demands a substantial amount of resources. The positioning of the terms "municipality," "mayor," "project," "support," and "water" reflects the presence of content related to institutional coordination, local projects, and the protection of natural resources, indicating that these elements are fundamental to the development and effective management of local projects.

In San José de Lourdes (Figure 3d), a clearly identifiable cluster of terms related to the strengthening of communal organization is observed, including "participation," "representatives," "local," "organization," "activities," "strengthen," "importance," and "need." The words "congress" and "assembly" appear together in the central area, highlighting the relevance of spaces where people gather to discuss and make decisions. Mining and the environment, by contrast, are located in a different sector.

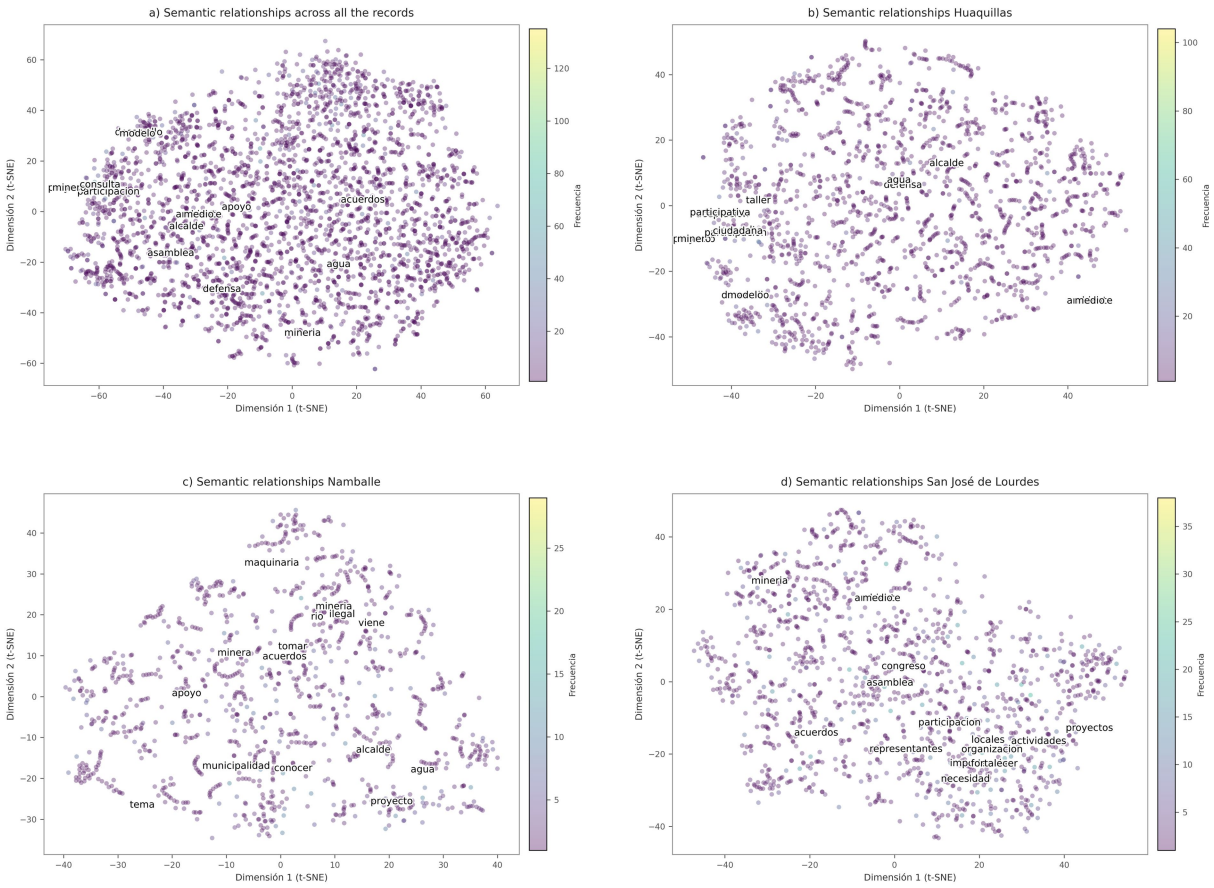


FIGURE 3
Two-dimensional semantic distribution of the terms present in the minutes of the Rondas Campesinas using Word2Vec and t-SNE: **(a)** Integrated corpus of the three localities; **(b)** Huaquillas; **(c)** Namballe; and **(d)** San José de Lourdes. Word2Vec embeddings and t-SNE projections were generated from the original Spanish-language corpus. The terms were retained in Spanish to preserve their contextual and semantic relationships and to avoid distortions potentially introduced through translation.

4 Discussion

TM techniques have become increasingly important for analyzing large document collections, particularly those containing unstructured or semi structured information, such as meeting minutes, reports, news articles, institutional records, and community documents (Cogburn et al., 2023; Hassani et al., 2020; Jadhav et al., 2023; Talabis et al., 2015). This study focused on the analysis of a textual corpus consisting of 74 meeting minutes from the rondas campesinas of Huaquillas, Namballe, and San Jose de Lourdes, located in the province of San Ignacio, department of Cajamarca, covering the period from 2012 to 2024. These documents are particularly relevant because the rondas campesinas not only carry out community surveillance and control activities, but also provide spaces for discussing local issues, making collective decisions, establishing community rules, and defending the territory.

The results indicate that the rondas campesinas are no longer confined to the community surveillance activities that originally led to their creation, revealing a significant transformation in their functions and responsibilities. Their minutes record activities such as the administration of justice, conflict resolution, territorial protection, institutional coordination, and participation in decisions affecting the community. This finding is consistent with previous research indicating that these organizations initially emerged in response to cattle theft but gradually expanded their activities to include community justice, social control, and the protection of collective interests (Chillihuani Ttito, 2020; Escalante, 2020; Nuñez Palomino, 1996; Piccoli, 2009).

A noteworthy finding is that the concerns identified are not expressed uniformly across the three localities. The observed differences suggest that each ronda campesina shapes its agenda according to the problems affecting its territory, its organizational experience, and the nature of its relationship with public authorities. In Huaquillas, the prominence of issues related to consultation and development models points to concern over the conditions under which mining projects are proposed and the extent to which local populations participate in decisions that may affect their resources. In Namballe, the relationship among mining, local authorities, and water resources reflects a context more closely associated with the regulation of extractive activities and the pursuit of institutional responses. In San José de Lourdes, the importance attributed to assemblies, representation, and organizational strengthening indicates that the community's capacity to respond relies primarily on its internal mechanisms of deliberation.

These differences show that rondas campesinas do not operate as homogeneous organizations, despite sharing principles of collective organization, territorial defense, and joint action. Their functioning is shaped by local conditions and specific historical processes, which are, in turn, influenced by social, economic, political, and cultural factors that vary across contexts. This

interpretation is consistent with Lázaro Aquino (2024), who argues that collective action in rural communities is constructed around territorial conflicts, shared needs, and particular relationships with state and economic actors. Accordingly, variations between localities should not be understood as inconsistencies, but rather as expressions of the organizational adaptability of the rondas campesinas.

The prominence of mining and environmental issues indicates that rondas campesinas conceive of territory as a social and collective space rather than merely as a source of economic resources. Discussions concerning water, land, the environment, and extractive projects show that territorial defense is closely linked to the continuity of productive activities, community security, and local living conditions. This finding is consistent with Shanee (2019), who identifies the rondas campesinas of northern Peru as key actors in environmental conservation and in the regulation of natural-resource use.

The TF-IDF analysis reveals that the discursive priorities of the rondas campesinas vary according to the characteristics of each territory. In Huaquillas, the relationship among development, consultation, and mining projects indicates that extractive activities are debated in connection with community participation and the preferred development model. In Namballe, the association among illegal mining, rivers, and public authorities reflects a stronger concern with regulating extractive activities and safeguarding water resources. In San José de Lourdes, the prominence of assemblies and organizational strengthening shows that collective deliberation constitutes a central mechanism for addressing local problems. These differences demonstrate that the rondas campesinas do not follow a uniform agenda, but respond to specific social, environmental, and institutional circumstances. This interpretation is consistent with the purpose of TF-IDF, which identifies the terms that distinguish one corpus from others beyond their mere frequency of occurrence (Kim and Gil, 2019; Qaiser and Ali, 2018).

The combined use of word clouds, Word2Vec, and t-SNE made it possible to examine both the predominant concepts and the relationships among them within rondero discourse. Word clouds provided an initial visual overview, whereas vector-based models enabled the identification of semantic associations derived from similar contexts of use (Heimerl et al., 2014; John et al., 2018; Lohmann et al., 2015; Mikolov et al., 2013; Pérez-Serrano et al., 2022). The observed relationships among mining, consultation, participation, water, authorities, and organization suggest that environmental defense is closely integrated with community decision-making and institutional coordination. Accordingly, issues concerning natural resources, territory, and organizational structure do not emerge in isolation, but as interconnected dimensions of the governance practices developed by the rondas campesinas.

Environmental pollution resulting from various productive and extractive activities is a subject of study for the social sciences due to emerging environmental social conflicts (Yearley, 2013), is corroborated by the first topic suggested from the minutes of the rondas campesinas in the three sites by LDA (Table 2). Highlighting the use of TM to study environmental conflicts, such as the regional ones in South Korea (Lee and Kim, 2020) and to measure the number of social conflicts related to the exploitation of non-renewable natural resources (Albrieu and Palazzo, 2020). These findings confirm that members of the rondas campesinas take on dynamic roles as leaders and initiators of conservation efforts in their local area, including the implementation of natural resource use control, extractive operations and environmental education (Shanee, 2019).

The findings generated through the different text-mining techniques proved to be mutually complementary, allowing the corpus to be examined from several analytical perspectives. Word clouds identified recurrent terms, TF-IDF distinguished the words that characterized each locality, LDA revealed underlying thematic structures, and Word2Vec combined with t-SNE represented semantic relationships among concepts used in similar contexts. Consequently, variations across the results should not be regarded as contradictions, but rather as distinct approaches to interpreting the discourse contained in the minutes. The integration of these techniques strengthens the analysis of large and unstructured corpora, particularly in research addressing social, environmental, and institutional issues (Chandran et al., 2025; Curiskis et al., 2020; Hofmann and Chisholm, 2016; Kataishi, 2026; Qamar and Raza, 2024).

From a sociological perspective, the findings indicate that the minutes do more than document agreements; they also reflect practices of communal governance, collective deliberation, coordination with public authorities, and the protection of shared resources. The presence of content related to assemblies, participation, agreements, mining, the environment, and authorities demonstrates that the rondas campesinas formulate organized responses to territorial problems. This dynamic is consistent with their recognition as organizations that perform functions of community justice and social coordination in areas where state presence remains limited or fragmented (Alarcón-Requejo, 2023; Nuñez Palomino, 1996; Piccoli, 2009). These findings may also be interpreted through the lens of legal pluralism, since the communal decisions recorded in the minutes coexist, interact, or come into tension with the state legal order within historical relationships shaped by territorial inequalities and persistent structures of power (Alarcón-Requejo, 2023; Casanova, 1965; Quijano, 2000).

TM techniques make it possible to identify lexical recurrences, patterns of association, and thematic clusters within a broad documentary corpus; however, these results do not, by themselves, constitute a complete sociological explanation. Their analytical value depends on the ability to situate this evidence within the historical, territorial, and institutional context in which the rondas campesinas operate. In this regard, recurring terms such as environment, mining, authorities, federation, and communities acquire deeper meaning when they are connected to the organizational trajectory of these institutions, their relationship with the State, the socio-environmental tensions present in the territory, and communal forms of deliberation, social control, and normative production. Therefore, the central contribution of this study does not lie solely in the application of computational tools to the analysis of community documents, but in demonstrating that these techniques can broaden the understanding of rural institutions of justice and governance when they are articulated with a sociological, contextual, critical, and intercultural interpretation.

Finally, the study has several limitations that should be considered when interpreting the findings. First, the analysis was based exclusively on 74 meeting minutes from three localities in the province of San Ignacio. Therefore, the results cannot be generalized to all rondas campesinas in Peru. Second, although TM techniques

made it possible to identify word frequencies, semantic associations, and latent topics, they cannot independently explain the motivations, internal tensions, or sociocultural meanings underlying the agreements documented in the records. During the transcription, cleaning, and tokenization of the corpus, certain discursive nuances characteristic of communal language may have been lost. Future research should therefore complement documentary analysis with interviews, observations of community assemblies, and comparative studies conducted in other regions. This would contribute to a deeper understanding of community justice, territorial governance, and legal pluralism in rural contexts.

5 Conclusion

The findings indicate that the priorities of the rondas campesinas vary according to the characteristics of each territory. In Huaquillas, the most prominent themes concerned development, consultation processes, and mining projects. In Namballe, illegal mining, water resources, and coordination with public authorities emerged as the principal concerns. In San José de Lourdes, assemblies and organizational strengthening were particularly salient. Across the full documentary corpus, these themes converged around development, environmental protection, community participation, and project management, demonstrating that the rondas campesinas link territorial conservation with processes of organization and collective decision-making.

The application of text-mining techniques enabled the identification of recurrent and distinctive terms, underlying thematic structures, and semantic relationships within the minutes. Word clouds, TF-IDF, LDA, and Word2Vec combined with t-SNE showed that environmental, mining-related, institutional, and organizational issues did not arise in isolation, but were interconnected within communal dynamics. These techniques proved particularly valuable for analyzing unstructured documents, identifying patterns, and achieving a more precise understanding of the concerns and forms of action developed by the rondas campesinas.

The study contributes to the sociology of law by demonstrating that the minutes not only record meetings and agreements, but also document and legitimize community-based normative practices. Through their assemblies, the rondas campesinas establish commitments, coordinate with public authorities, and organize responses to environmental, economic, and institutional problems in contexts where communal justice coexists, engages with, or comes into conflict with the state legal order. Future research could complement these findings through interviews, observations of assemblies, and comparative studies conducted in other regions, thereby deepening understanding of communal justice, territorial governance, and legal pluralism in rural areas of Peru.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

LQ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. RD: Conceptualization, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. NA: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft, Data curation, Validation. FF: Conceptualization, Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. JM: Conceptualization, Formal analysis, Investigation, Validation, Visualization, Writing – original draft. OC: Conceptualization, Formal analysis, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing. SS: Investigation, Validation, Visualization, Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was funded by the Vice-Rectorate for Research of the Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM), Peru.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acuña, J. K. (2019). The common in times of war. The Rondas Campesinas and the Maoist movement in Peru. *Encuentros* 9, 55–68. Available online at: https://encuentros.unermb.web.ve/index.php/encuentros/article/view/47

Alaminos-Fernández, A. (2003) *Introducción a la minería de texto y análisis de sentimiento con R* Universidad de Alicante. Available online at: https://rua.ua.es/server/api/core/bitstreams/84d03042-ccb4-40f8-a4bb-0414048ed74f/content (accessed May 6, 2026).

Alarcón-Requejo, G. (2023). Challenges of coordination between communal justice of the peasant patrols and ordinary justice in Peru: a critical review of plenary agreement no. 1-2009/CJ-116. *Juridicas* 20, 75–93. doi: 10.17151/jurid.2023.20.1.4

Albrieu, R., and Palazzo, G. (2020). *Mapping Social Conflicts in Natural Resources: A Text Mining Study of Extractive Activities*. Available online at: https://hdl.handle.net/11362/46586 (accessed April 6, 2026).

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022. doi: 10.5555/944919.944937

Brisman, A., South, N., and White, R. (2016). *Environmental Crime and Social Conflict: Contemporary and Emerging Issues*. Routledge. Available online at: https://books.google.com.mx/books?hl=es&lr=&id=orG1CwAAQBAJ&oi=fnd&pg=PP1&dq=Environmental+crime+and+social+conflict:+contemporary+and+emerging&ots=mJWUZUPN8U&sig=_YnCppzes6iZo-LBRaTrjnJrCwA (accessed April 6, 2026).

Casanova, P. G. (1965). Internal colonialism and national development. *Stud. Comp. Int. Dev.* 1, 27–37. doi: 10.1007/BF02800542

Chandran, N. V., Anoop, V. S., and Asharaf, S. (2025). Topic-weighted kernels: text kernels integrating topic weights and deep word embeddings for semantic text analytics. *IEEE Access* 13, 77918–77930. doi: 10.1109/ACCESS.2025.3565816

Chillihuani Ttito, V. (2020). Las rondas campesinas del Perú una alternativa de justicia en las zonas rurales alto andinas, el caso de Ocongate un distrito rural del departamento del cusco 1992-2011. *Horizonte Cienc.* 10, 55–74. doi: 10.26490/uncp.horizonteciencia.2020.18.429

Choi, C., Lee, J., Machado, J., and Kim, G. (2022). Big-data-based text mining and social network analysis of landscape response to future environmental change. *Land* 11:2183. doi: 10.3390/land11122183

Cogburn, D., Hine, M., and Yoon, V. (2023) Introduction to the Minitrack on text mining and analytics. *Hawaii International Conference on System Sciences*. Honolulu, HI, USA: ScholarSpace, University of Hawai'i at Mānoa.

Curiskis, S. A., Drake, B., Osborn, T. R., and Kennedy, P. J. (2020). An evaluation of document clustering and topic modelling in two online social networks: twitter and Reddit. *Inf. Process. Manag.* 57:102034. doi: 10.1016/j.ipm.2019.04.002

Davis, D. E. (1999). The power of distance: re-theorizing social movements in Latin America. *Theory Soc.* 28, 585–638. doi: 10.1023/A:1007026620394

de Boer, M. H. T., Bakker, B. J., Boertjes, E., Wilmer, M., Raaijmakers, S., and van der Kleij, R. (2019). Text Mining in Cybersecurity: exploring threats and opportunities. *Multimodal Technologies Interaction* 3:62. doi: 10.3390/mti3030062

del Castillo Zayas, Y. M., and Leiva Mederos, A. A. (2007). La minería de texto: Perspectiva metodológica para la realización de resúmenes documentales. *ACIMED* 15. Available online at: https://agora.edu.es/servlet/articulo?codigo=2285589

Desai, A., Zumbo, A., Giordano, M., Morandini, P., Laino, M. E., Azzolini, E., et al. (2022). Word2vec word embedding-based artificial intelligence model in the triage of patients with suspected diagnosis of major ischemic stroke: a feasibility study. *Int. J. Environ. Res. Public Health* 19:15295. doi: 10.3390/ijerph192215295

Doerr, N., Mattoni, A., and Teune, S. (2013). Toward a visual analysis of social movements, conflict, and political mobilization. En *Advances in the Visual Analysis of Social Movements* (pp. xi–xxvi). Emerald Group Publishing Limited. Available online at: https://www.emerald.com/insight/content/doi/10.1108/S0163-786X(2013)0000035004/full/html (accessed April 7, 2026).

Dong, M., Lu, J., Wang, G., Zheng, X., and Kiritsis, D. (2022). Model-based systems engineering papers analysis based on word cloud visualization. *2022 IEEE International Systems Conference (SysCon)*. Montreal, QC, Canada: IEEE.

Edelman, M. (2001). Social movements: changing paradigms and forms of politics. *Annu. Rev. Anthropol.* 30, 285–317. doi: 10.1146/annurev.anthro.30.1.285

Elgendy, N., and Elragal, A. (2014). "Big data analytics: a literature review paper," in *Advances in Data Mining. Applications and Theoretical Aspects*, ed. P. Perner (Cham: Springer International Publishing), 214–227.

Escalante, E. E. (2020). Night watchers and terrorists. *J. Econ. Behav. Organ.* 171, 116–131. doi: 10.1016/j.jebo.2019.12.025

Feldman, R., and Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. doi: 10.1017/CBO9780511546914

Fones, D., and Fones, C. S. L. (2025). AI-powered chatbots in psychiatry: a critical evaluation of ChatGPT's sleep aid guidance. *Int. J. Neuropsychopharmacol.* 28, i245–i246. doi: 10.1093/ijnp/pyae059.431

Giugni, M. (1999). "How social movements matter: past research, present problems, future developments," in *How social Movements Matter*, (Minneapolis, MN, USA: University of Minnesota Press), xiii–xxxiii.

Grotenhuis, Z., Mosteiro, P. J., and Leeuwenberg, A. M. (2024). Modest performance of text mining to extract health outcomes may be almost sufficient for high-quality prognostic model development. *Comput. Biol. Med.* 170:108014. doi: 10.1016/j.compbiomed.2024.108014

Gyódi, K., Nawaro, Ł., Paliński, M., and Wilamowski, M. (2023). Informing policy with text mining: technological change and social challenges. *Qual. Quant.* 57, 933–954. doi: 10.1007/s11135-022-01378-w

Hamel, P., and Maheu, L. (2001). "Beyond new social movements: social conflicts and institutions," in *The Blackwell Companion to Political Sociology*, (), 261–270.

Han, K., Yeom, J., and Chung, K. (2024). Identifying emerging issues in the seafood industry based on a text mining approach. *Appl. Sci.* 14:1820. doi: 10.3390/app14051820

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., and Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data Cogn. Comput.* 4:1. doi: 10.3390/bdcc4010001

Heimerl, F., Lohmann, S., Lange, S., and Ertl, T. (2014) Word cloud explorer: text analytics based on word clouds *2014 47th Hawaii International Conference on System Sciences*. Waikoloa, HI, USA: IEEE, 1833–1842.

M. Hofmann and A. Chisholm (2016). *Text Mining and Visualization: Case Studies Using Open-Source Tools*. Taylor Francis, NY: Chapman and Hall/CRC. doi: 10.1201/b19007

Inclán, M. (2018). Latin America, a continent in movement but where to? A review of social movements' studies in the region. *Annu. Rev. Sociol.* 44, 535–551. doi: 10.1146/annurev-soc-073117-041043

Jadhav, A., Jagtap, P., Gurav, S., Jadhav, S., Jadhav, N., and Akkalkot, A. (2023). A survey on text mining—techniques, application. *Int. J. Scientific Res. Comput. Sci. Eng. Inform. Technol.* 9, 338–343. doi: 10.32628/CSEIT2390391

Jayashankar, S., and Sridaran, R. (2017). Superlative model using word cloud for short answers evaluation in eLearning. *Educ. Inf. Technol.* 22, 2383–2402. doi: 10.1007/s10639-016-9547-0

John, M., Marbach, E., Lohmann, S., Heimerl, F., and Ertl, T. (2018). "MultiCloud: interactive word cloud visualization for the analysis of multiple texts," in *Proceedings of Graphics Interface 2018*, (Toronto: ACM Digital Library Board).

Kataishi, R. (2026). Enhancing retrieval-augmented generation with topic-enriched embeddings: a hybrid approach integrating traditional NLP techniques. *Natu. Lang. Process. J.* 14:100200. doi: 10.1016/j.nlp.2026.100200

Kim, S.-W., and Gil, J.-M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Hum. Cent. Comput. Inf. Sci.* 9:30. doi: 10.1186/s13673-019-0192-7

Kitchin, R. (2021). *Data Lives: How Data Are Made and Shape Our World*. Policy Press. doi: 10.1332/policypress/9781529215144.001.0001

Lázaro Aquino, T. G. (2024). Las comunidades campesinas en el Perú: Organización, conflictos y acción colectiva, Junín, 1959-2017. *Desde el Sur* 16, 1–27. doi: 10.21142/des-1602-2024-0025

Lee, J., and Kim, D. (2020). Mapping environmental conflicts using spatial text mining. *Land* 9:287. doi: 10.3390/land9090287

Lohmann, S., Heimerl, F., Bopp, F., Burch, M., and Ertl, T. (2015) Concentri cloud: word cloud visualization for multiple text documents *2015 19th International Conference on Information Visualisation*. Barcelona, Spain: IEEE, 114–120.

Lopez-Flores, P.-C. (2024). Territorial re-existences of indigenous movements in the Andean-Amazon region. *Anduli* 25, 1–20. doi: 10.12795/anduli.2024.i25.01

Manning, C. D., Raghavan, P., and Schütze, H. (2009) *An Introduction to Information Retrieval* Cambridge University Press Cambridge, England. Available online at: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf (accessed May 6, 2026).

Markellos, K. (2009). "Text Mining for Business Intelligence," in *Encyclopedia of Data Warehousing and Mining*. 2nd ed (NY: IGI Global Scientific Publishing), 1947–1956. doi: 10.4018/978-1-60566-010-3.ch298

Melucci, A. (1984). An end to social movements? Introductory paper to the sessions on "new movements and change in organizational forms". *Soc. Sci. Inf.* 23, 819–835. doi: 10.1177/053901884023004008

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). *Distributed Representations of Words and Phrases and Their Compositionality. Advances in Neural Information Processing Systems*. Available online at: https://proceedings.neurips.cc/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html (accessed April 7, 2026).

Nuñez Palomino, G. (1996). The rise of the Rondas Campesinas in Peru. *J. Leg. Plural. Unoff. Law* 28, 111–123. doi: 10.1080/07329113.1996.10756468

Olsher, D. J. (2015). New artificial intelligence tools for deep conflict resolution and humanitarian response. *Procedia Eng.* 107, 282–292. doi: 10.1016/j.proeng.2015.06.083

Palazzo, G. (2017). Minería de Datos, Conflictos Sociales y Explotación de Recursos Naturales: Un Aporte Metodológico. *Cuadernos de Economía y Administración* 4. Available online at: https://www.revistas.espol.edu.ec/index.php/compendium/es/article/view/179/165

Patel, P. (2017). *Big Data: Challenges and Opportunities*.

Percha, B. (2021). Modern clinical text mining: a guide and review. *Annu. Rev. Biomed. Data Sci.* 4, 165–187. doi: 10.1146/annurev-biodatasci-030421-030931

Pérez, I. R., and Reyes, Y. M. (2022). Aplicación de la minería de texto como técnica de análisis para la identificación de enfoques temáticos en un grupo de investigación. *Investigación e Innovación en Ingenierias* 10, 178–195. doi: 10.17081/invinno.10.2.5907

Pérez-Serrano, M., Nogueroles-López, M., and Dunabeitia, J. A. (2022). Effects of semantic clustering and repetition on incidental vocabulary learning. *Front. Psychol.* 13:997951. doi: 10.3389/fpsyg.2022.997951

Piccoli, E. (2009). Las rondas campesinas y su reconocimiento estatal, dificultades y contradicciones de un encuentro: Un enfoque antropológico sobre el caso de Cajamarca, Perú. *Nueva Antropol.* 22, 93–113.

Ptaszek, G., Yuskiv, B., and Khomych, S. (2024). War on frames: text mining of conflict in Russian and Ukrainian news agency coverage on telegram during the Russian invasion of Ukraine in 2022. *Media War Confl.* 17, 41–61. doi: 10.1177/17506352231166327

Qaiser, S., and Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *Int. J. Comput. Appl.* 181, 25–29. doi: 10.5120/ijca2018917395

Qamar, U., and Raza, M. S. (2024). "Text mining applications," in *Applied Text Mining*, eds. U. Qamar and M. S. Raza (Cham: Springer Nature Switzerland), 53–81. doi: 10.1007/978-3-031-51917-8_3

Quijano, A. (2000). Coloniality of power and eurocentrism in Latin America. *Int. Sociol.* 15, 215–232. doi: 10.1177/0268580900015002005

Rezaev, A. V., and Tregubova, N. D. (2018). Are sociologists ready for 'artificial sociality'? Current issues and future prospects for studying artificial intelligence in the social sciences. *Мониторинг общественного мнения* 5 (147), 91–108.

Sakthi Vel, S. (2021) Pre-processing techniques of text mining using computational linguistics and Python libraries *2021 International Conference on Artificial Intelligence and Smart Systems* (Coimbatore, India: IEEE) 879–884.

Salton, G. (1983). *Introduction to Modern Information Retrieval.* McGrawHill Book Co. Available online at: https://cir.nii.ac.jp/crid/1574231873785747328 (accessed April 7, 2026).

Sawitri, D. (2024). Big data challenges and opportunities in the development of digital technology. *J. Info Sains* 14:Article 02. doi: 10.54209/infosains.v14i02.4619

Shanee, N. (2019). Reclaim conservation: conservation discourse and initiatives of the Rondas Campesinas, north-eastern Peru. *Conserv. Soc.* 17, 270–282. doi: 10.4103/cs.cs_18_6

Singh, T., Kumari, M., Pal, T. L., and Chauhan, A. (2017). Current trends in text mining for social media. *Int. J. Grid Distrib. Comput.* 10, 11–28. doi: 10.14257/ijgdc.2017.10.6.02

Soliman, M., Fatnassi, T., Elgammal, I., and Figueiredo, R. (2023). Exploring the major trends and emerging themes of artificial intelligence in the scientific leading journals amidst the COVID-19 era. *Big Data Cogn. Comput.* 7:12. doi: 10.3390/bdcc7010012

Talabis, M. R. M., McPherson, R., Miyamoto, I., Martin, J. L., and Kaye, D. (2015). "Security and text mining," in *Information Security Analytics*, (NY: Syngress is an Imprint of Elsevier), 123–150.

Talib, R., Hanif, M. K., Ayesha, S., and Fatima, F. (2016). Text mining: techniques, applications and issues. *Int. J. Adv. Comput. Sci. Appl.* 7:11. doi: 10.14569/IJACSA.2016.071153

Taskesen, E., and Reinders, M. J. T. (2016). 2D representation of transcriptomes by t-SNE exposes relatedness between human tissues. *PLoS One* 11:e0149853. doi: 10.1371/journal.pone.0149853

Touraine, A. (1985). An introduction to the study of social movements. *Soc. Res.*, 749–787.

Wiggins, C., and Jones, M. L. (2023). *How Data Happened: A History from the Age of Reason to the Age of Algorithms*. New York, NY, USA: W. W. Norton & Company.

Xiaojun, Z. (2011). Michael W. Berry and Jacob Kogan (eds.): text mining: applications and theory. *Inf. Retr.* 14, 208–211. doi: 10.1007/s10791-010-9153-5

Yamin, M. (2019). Information technologies of 21st century and their impact on the society. *Int. J. Inf. Technol.* 11, 759–766. doi: 10.1007/s41870-019-00355-1

Yearley, S. (2013). Social movements and environmental change. In *Social Theory and the Global Environment* (pp. 150–168). Routledge. Available online at: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203427903-7/social-movements-environmental-change-steven-yearley (accessed April 7, 2026).

Yoon, J., Han, S., Lee, Y., and Hwang, H. (2023). Text mining analysis of ESG management reports in South Korea: comparison with sustainable development goals. *SAGE Open* 13:21582440231202896. doi: 10.1177/21582440231202896

Yun-tao, Z., Ling, G., and Yong-cheng, W. (2005). An improved TF-IDF approach for text classification. *J. Zhejiang Univ. A* 6, 49–55. doi: 10.1631/BF02842477

Zanini, N., and Dhawan, V. (2015). *Text Mining: An Introduction to Theory and Some Applications*. Cambridge, UK: Cambridge University Press & Assessment.