


RESEARCH

Open Access



Finding the context indigenous innovation in village enterprise knowledge structure: a topic modeling

Retno Kusumastuti¹, Mesnan Silalahi^{2*} , Anugerah Yuka Asmara², Ria Hardiyati² and Vishnu Juwono¹

*Correspondence:

mesnans@yahoo.com

² Research Center for STI

Policy & Management,

Indonesian Institute

of Sciences, Jakarta,

Indonesia

Full list of author information
is available at the end of the
article

Abstract

Indigenous people have deep local knowledge of environmental sustainability and natural resource utilization, which are sources of innovations that often are drivers for economic growth in rural areas. This study explores the knowledge structure of indigenous innovation in village enterprises through content analysis of research publications. The resulting knowledge structure can be used to set up a roadmap for the studies on village enterprise and in a broader context to build metadata as a foundation for an evaluation system of village enterprise. The authors deploy topic modeling and co-word analyses to scrutinize 775 village enterprise research articles from the Scopus database and 665 paper from ScienceDirect. In the topic modeling, topic models village enterprises are setup. The topics found are local ownership (such as market and property), land, services (housing, health care), economy and public policy, financial service micro-credit, environmental pollution control, local business sustainability, social entrepreneurship, and household income, bioenergy based electrification, and bumdes management. Four sectors of the natural resource-based indigenous economy were identified: traditional food production, bio-energy for fuel and electricity, agriculture, and tourism. The topic models are used to comprehend knowledge structure in the village enterprises, whereby the focus is to uncover the context of indigenous village enterprise and its states of the art.

Keywords: Indigenous people, Village enterprise, Knowledge structure, Topic modeling analysis

Introduction

The distribution of indigenous people around the world is quite significant. They occupy about 22% of the earth's surface (ILO, 2017), comprising less than 5% of the world's population, and protect 80% of global biodiversity (Gleb Raygorodetsky, 2018). Unfortunately, these community groups' rights are often marginalized. Indigenous people have deep local knowledge related to environmental sustainability and natural resources called indigenous knowledge (Capel, 2014). These indigenous knowledge supported with technology are sources of innovations that often are drivers for economic growth in rural areas. The issues on village enterprise are increasingly important, especially for Indonesia, whereby a lot of money massively has been allocated for village funds in the last

5 years (Gatra, 2019; Undang-Undang Republik Indonesia Nomor 6 Tahun 2014 Tentang Desa, 2014). In Indonesia, this large amount of money is the enabler to the sustainability of village enterprises well known as BUMDes, the pillars of economic activities in the village that function as social institutions and commercial institutions. BUMDes as a social institution favors the community's interests through its contribution to social services provision. In the same context as a commercial institution, it aims to seek profits by offering local resources (products and services).

The context of indigenous people has become a concern of the United Nations. It has ratified several clauses related to indigenous people's rights, especially in economic development and economic independence. Therefore, many studies have been conducted on village enterprises on various aspects and approaches to better understand the complexity. Capel (2014) and Howell (2018) confirm that indigenous innovation is one strategy for driving the economy and business sustainability. According to resource based perspective, local wisdom and indigenous knowledge in one area could be a competitive advantage, since it contains valuable characteristics, rare and hard to imitate. The framework used to analyze the competitive advantage is VRIO (Value, Rare, Imitability, Organization) analysis framework (Barney & Clark, 2007).

The issue of strengthening the village enterprise is important considering that many studies have shown that the village enterprise is the backbone of the economy in marginalized villages and traditional villages. Most of the village enterprises are micro, small, and medium in size managed by indigenous peoples. Indigenous knowledge plays an important role in the development of several sectors, including forestry, agriculture, food, medicine, and tourism. Studies in several regions show that traditional businesses that rely on indigenous knowledge in their production processes are proven to be in line with government efforts in sustainability and environment protection. Indigenous knowledge produced based on local resources, technologies, and local culture (Boon & Hens, 2007; Dewalt, 1994) is environmentally sustainable in many cases and an important source of rural resilience and, to some extent, strengthens self-determination.

Topic modeling is an unsupervised classification method used for documents, with the purpose of detecting topics from the document sets (corpus). Topic modeling analysis is now attracting the studies on social sciences because of the new possibilities from analyzing scientific text documents, or using social media data to understand the prevalent topics and their evolution. Related studies using topic modeling are such as on identifying social entrepreneurship strategies (Chandra et al., 2016), and on understanding corporate social responsibility (Chae & Park, 2018). The topic modeling, could represent a valid alternative in identifying interdisciplinary fields directly from the textual content of scientific papers. Chandra et al. (2016) classified 39 strategies in social entrepreneurship (SE) using structural topic modeling, and then reduced it to six meta-SE strategies. Four new meta-SE strategies were identified that have been overlooked in prior SE research. Kao and Luarn (2020) explored the ideas of social enterprises through a bottom-up approach using twitter data, and classifies four dimensions (strategy, impact, business and people) and six indicators (social, opportunity, change, enterprise, network and team) to establish a conceptual framework of social enterprises. Latent Dirichlet Allocation (LDA) is a topic modeling technique to extract topics from text documents, where each topic is a mixture over words and each document is a mixture over topics. To find

the convenient number of topics, a few topic models are trained using different numbers of topics and evaluated them with measures of log-likelihood (Wallach et al., 2009), the perplexity (Blei et al., 2003), or semantic coherence of the topics (Mimno et al., 2011; Newman et al., 2009).

This paper attempted to find the structures of knowledge in the village enterprise research' publications and find the linkage to indigenous innovation. There are not many studies that take a research locus on indigenous village enterprise, so it is necessary to map the global empirical studies development. To this end, this study elaborates topic modeling to provide more insight into the knowledge structure of village enterprises based on publications from the Scopus database and Science Direct, with the end goal to build metadata on village enterprises. Because of its diversity, metadata is needed for better evaluation of the development of a village enterprise. The rationale for the development of a metadata is that the diverse indigenous knowledge generated from the village and village enterprises will certainly grow into indigenous innovation practices that needs an information system to manage the knowledges effectively.

Studies of indigenous knowledge and indigenous innovation

Indigenous knowledge studies have been used in various innovation studies (Appelbaum et al., 2016; Baskaran & Mehta, 2016; Capel, 2014; Huang et al., 2018; Jauhainen & Hooli, 2017; Mika et al., 2017). Empirical evidence in most of Africa's population in the south of Sahara showed that indigenous knowledge forms the basic foundation of their innovation and invention (Ezeanya-Esiobu, 2019). The existing indigenous knowledge studies not only focus on increasing the economic development of a region but also on how an area's economy progress while maintaining the culture and the condition of the existing natural resources. Therefore, indigenous knowledge also functions in solving social issues in the region (Baskaran & Mehta, 2016; Capel, 2014; Padilla-Meléndez & Ciruela-Lorenzo, 2018). Indigenous knowledge studies are then expanded to define indigenous communities (Blackman & Veit, 2018; Curry et al., 2016; Karanasios & Parker, 2018; Makondo & Thomas, 2018; Padilla-Meléndez & Ciruela-Lorenzo, 2018). Besides, there are studies emphasizing the importance of local-traditional values and culture in the field of entrepreneurship and starts-up (Capel, 2014; Curry et al., 2016; Padilla-Meléndez & Ciruela-Lorenzo, 2018), also in the field of new and renewable energy which discusses the attitude of local communities to accept or reject foreign technology in their area. The issues are carbon emissions, renewable energy, land, water, and forest, and how local communities and governments manage energy and technology independently in their regions (Karanasios & Parker, 2018). These last issues then provide a place for the contribution of indigenous knowledge to solving global warming in various countries (Makondo & Thomas, 2018), including how the role of indigenous knowledge in maintaining forests in Latin American countries (Ecuador, Colombia, Brazil, and Bolivia) to reduce the increase in world carbon emissions (Blackman & Veit, 2018). In Kenya, in a rural area of Turkana, the use of indigenous knowledge by considering local culture and values tends to maintain local resources and preserved natural environments while stimulating economic development based on natural and local tourism. The challenge for local communities is to adopt new knowledge and technology to be aligned with existing indigenous knowledge (Ng'Asike & Swadener, 2015).

Various studies have shown that government involvement is important in encouraging indigenous knowledge and indigenous innovation. Likewise, the existing institutional patterns, whether formed by the government or collective agreement, influence or maintain indigenous knowledge and indigenous innovation practices in an area (Li-Ying & Wang, 2015; Yang et al., 2014; Zhao et al., 2015). Adopting the success of indigenous practices as has been done by several studies is not necessarily direct. Each region and society has different local characteristics (thus many attributes), historical factors that have existed in the region, and various other factors that are very complex in nature, which become a consideration for not applying and adopting indigenous practices from one place to another. For this reason, it is important to carry out indigenous learning intensely and thoroughly with local adaptive context standards. Conducting research driven by local characteristics with many cases and bringing out each community's uniqueness in an area is important before indigenous practices elsewhere are implemented in new places (Nelson-Barber & Johnson, 2016). For example, a study from Padilla-Meléndez and Ciruela-Lorenzo (2018), which discusses female indigenous entrepreneurs, found that entrepreneurial practice in indigenous women's groups is not only a factor of common interests and goals of these women's groups but also the existence of social capital ties and also individual motives of the group. Although indigenous communities are important, indigenous factors that encourage economic entrepreneurship development also need to be considered. Even though local culture contributes to the development of female indigenous entrepreneurs, the values in it need to be corrected because of these groups' individual motives, which become the bonds of cooperation within these indigenous groups.

Materials and methods

The global research publications related to village enterprise were searched in the Scopus database and Science Direct. The phrase "village enterprise", "rural enterprise", and "bumdes" were searched in the topic field (title/abstract/keyword) in the database. The examined publications are until October 2020, using search script as follows: TITLE-ABS-KEY ("village enterpr*" OR "bumdes" OR "rural enterpr*" OR "rural busine*" OR "village fund*" OR "village owned enterprise*"). Titles and abstracts of all publications were carefully considered for relevance to village enterprise. Publications that were not related to village enterprise, or duplicate publications, were excluded from the results of the search query. As a result, a total of 1440 publications were retained for the next stage. Before analysis, the common preprocessing of the text consists of three steps: tokenization, stop word removal, and stemming. Tokenizing is the process of dividing the content of each text into a sequence of character strings called tokens. This will generate a token consisting of a single word before finally building the word vector. Stop word removal means eliminating the filler words that are often used, or often called stop word, which does not add value to the analysis. Stemming involves removing word endings to reduce vocabulary size, and words are returned to the root word (Porter, 2006). Depending on the objectives of a study, stemming can mean better results or an increase in errors (Manning et al., 2008). In this study, stemming is not used to get a more straightforward interpretation of the results.

Topic modeling

LDA is the most widely used topic modeling originally developed by Blei et al., (2003), which introduced the initial Dirichlet distribution to the topic-and-word document distribution, encoding the intuition that the document covers a number of topics and that topics use a set of words. This model can reveal the main topic of a corpus that can potentially be used to build knowledge structures in a domain of interest. This quantitative method does not offer the depth of contextual understanding that qualitative methods do.

From a data set (a collection of documents or a corpus), LDA backtracks and determines what topics will make up the document. The corpus is represented as a matrix of terms in a document (DTM), which is generally very rare (sparse matrix). Reducing the dimensions of the matrix can improve the topic modeling results. For this purpose, preprocessing is necessary so that syntactically close words can be included in just one basic term. Figure 1 shows a graphical representation model of the LDA using plate notation, which illustrates the dependencies between model parameters. The plate box represents the text. The outer plate represents the document, while the inner plate represents the topic choices and repetitive words in the document.

The total probability of the corpus can be calculated by the formula:

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d \tag{1}$$

The LDA model is represented as a probabilistic graphical model in the diagram above. There are three levels to the LDA representation. M represents the total documents in the corpus, while N represents the number of words in a document. Parameters α and β are corpus level parameters; it is assumed that the sample is taken once in the process of producing the corpus. α is the parameter of the initial Dirichlet on the per-document topic distribution, β is the parameter of the initial Dirichlet on the word-by-topic distribution. The variable θ_d is a document-level variable that represents the topic distribution for document d , which is taken once per document. Finally, the variables z_{dn} (the topic for the n th word in document d) and w_{dn} (the specific word) are word-level variables and are taken once for each word in each document. Topic model selection is carried out is based on the minimum perplexity value, which is defined as

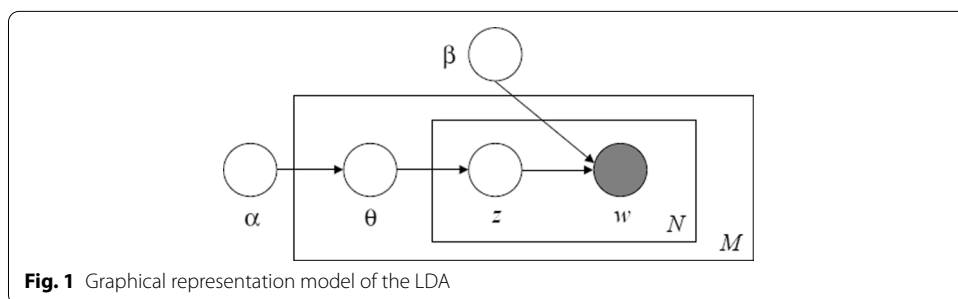


Fig. 1 Graphical representation model of the LDA

$$\text{Perplexity}(w) = \exp \left\{ - \frac{\log(p(\varphi))}{\sum_{d=1}^D \sum_{j=1}^V n^{(jd)}} \right\} \quad (2)$$

In information theory, perplexity is a measure of how well a probability model predicts a sample to determine the statistical goodness of fit of a topic model (Blei et al., 2003). It can be used to compare probability models. A low perplexity value indicates a good probability distribution in predicting the sample which would give results that making it easier to interpret. Chang et al. (2009) showed models which achieve better predictive perplexity often have less interpretable latent spaces.

The pursuit toward a better method to tackle the interpretability issues of a topic model and the topic size determination was also directed to evaluating the semantic coherence of the topic models. Semantic coherence is a measure of the co-occurrence of highly probable words in a topic and has been shown to correlate with expert judgments of topic quality (Mimno et al., 2011). Newman et al. (2009) proposed a scoring model using pointwise mutual information (PMI) with external data to evaluate the semantic coherence of the topic models. The model was proposed from the facts that a topic has some odd-words in the list of top-ten words. The coherence of topics is calculated based on a sliding window with the size 10 and the pointwise mutual information (PMI) of all word pairs of the given top words. The coherence is the result from the arithmetic mean of the PMI values. The PMI of all a given word pair (w_i, w_j) is calculated as

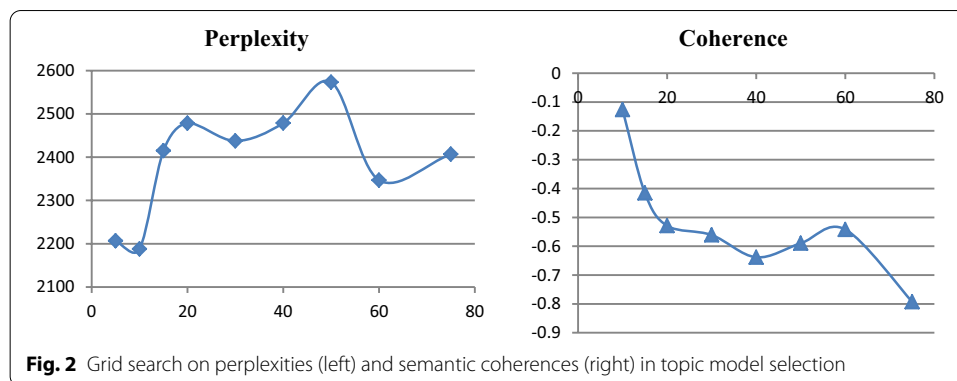
$$\text{PMI}(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}. \quad (3)$$

In this paper the selection of the models are based on the triangulation of the perplexity and the semantic coherence following Newman et al. (2009). The selected individual topics are evaluated and compared on its interpretability and the theoretical concept (Bonilla & Grimmer, 2013; Maier et al., 2018). The dynamics of the topics are calculated using labeled LDA (Ramage et al., 2009) which can classify a document with multiple labels, and is useful to study how the topics evolve over time.

Results and discussion

In the selection process of the topic model a grid search on several other topic sizes were carried out, namely, $k=5, 10, 15, 20, 30, 50, 60,$ and 75 . This was carried out to gain knowledge about the granularity of large topics. From Fig. 2 (left), the optimal searched values of the perplexity were selected at three points, namely, the topic model with topic sizes 15, 30 and 60. These three topic models are supported by results from the coherence values calculation (right) which are then used as a basis for interpreting the knowledge structure on village enterprise.

In this study, the focus is placed on finding concepts in the text (publications) related to the indigenous rural enterprise. These concepts may be hidden in the large volume of research documents. For this reason, efforts were made to find the concept by modeling in various ways, such as modeling with several topic sizes (topic size— k). Modeling also considers finding these hidden concepts using a large enough top-term. The top term count is the number of main words that make up a topic, which is usually decided according to need. If the problem statement focuses on extracting



a theme or concept, it is advisable to choose a higher number; if the problem statement focuses on extracting a feature or term, a low number is recommended. Table 1 provides a topic model with a value of $k=15$ with meaningful topics, such as: traditional food production, local ownership (such as market and property), land, public services (housing, health care, retirement), economic policy, financial micro-credit, environment pollution control, employment, local business sustainability, electricity, women (gender) and household income, bumdes management, and public policy. Using a topic model with $k=30$, some subtopics related to topics in the topic model with $k=15$ can be constructed. For each topic, the intensity is calculated by summing the probability of every topic in each of the documents.

In the topic of traditional food production (A00), terms, such as management, supply chain, quality and cost are aspects of the topic. The topic of local ownership includes aspects, such as market, government, reform, privatization, property, rights, collective, institutional firms, political, control, and governance. Topic public/community services (A02) is covering aspects of housing, healthcare and retirement. Topic 03 is about the village economy, where the aspects are market in the agricultural sector, policy, growth, employment, population and labor. Topic 06 is a topic on electricity, whereby generation from biomass and biogas is dominant in a rural area's renewable energy generation. This topic has subtopics, as shown in Table 2, namely, the topic of electricity supply and demand system for households, business, and community (B17), biomass-based renewable energy technology for electricity (B03), and biogas production (B22). In the business topic (A07), the issues are access, sustainability, and resilience related to the environment. Topic A08 on the environment is closely related to water pollution and the river. Topic A12 on BUMDes management considers aspects of the community, financial system, technology support, planning, sustainability, challenges, and policy. Topic A05 on household income is related to women, education, and inequality. In Table 2 can be found subtopics of business (A08) from the topic model in Table 1, which are related to innovation (B18) and entrepreneurship (B24). The context "indigenous" was found related to sustainability as in topic B13. The terms in sequence reflect increasing weighted value: forest, landscape, ecological, natural resources, tourism, social, environmental, planning, conservation, species, cultural, land, and biodiversity. The interpretation would be that forest, landscape, ecological natural resources, biodiversity conservation, and embedded

Table 1 Topic model of village enterprise publication with topic size $k = 15$

Topic no	Top terms	Topic intensity
A00	Production, analysis, food, farmers, management, supply, industry, chain, products, level, model, models, application, quality, costs, method, information, project, enterprise, traditional	61.84
A01	Local, township, ownership, tves, market, government, reform, private, governments, Chinese, property, privatization, rights, collective, institutional, firms, political, control, enterprise, governance	110.01
A02	Housing, service, care, health, integrated, retirement, services, villages, community, people, provision, sector, living, Australia, private, public, homes, residents, industry, forms	15.68
A03	Economy, market, policy, policies, countries, agricultural, sector, growth, agriculture, industry, government, employment, system, structure, population, changes, issues, labour, political, Chinese	130.90
A04	Credit, sector, labor, employment, financial, urban, growth, informal, institutions, effects, investment, migration, workers, household, nonfarm, finance, activities, households, income, evidence	88.65
A05	Income, women, poverty, households, household, livelihood, education, family, activities, poor, villages, access, inequality, farm, impact, reduction, diversification, factors, alleviation, province	62.20
A06	Electricity, biomass, system, systems, India, biogas, production, consumption, generation, solar, khadi, electrification, potential, renewable, cost, plant, performance, demand, coal, total	76.74
A07	Business, local, access, sustainability, sustainable, smes, services, businesses, analysis, environmental, firms, information, resilience, developing, approach, technology, countries, marketing, supply, performance	84.93
A08	Environmental, water, pollution, environment, river, carbon, health, protection, industrial, emissions, evaluation, weight, control, quality, increased, system, factors, measures, concentration, waste	50.70
A09	Social, local, enterprise, business, role, entrepreneurship, support, literature, factors, understanding, findings, sector, framework, analysis, networks, tourism, context, processes, capital, approach	290.51
A10	Growth, tves, Chinese, industrial, efficiency, economy, township, investment, period, industries, sector, industry, provinces, capital, regional, production, analysis, output, productivity, level	138.32
A11	Land, urban, urbanization, spatial, industrial, regional, cities, social, landscape, town, factors, pattern, construction, region, model, process, distribution, city, towns, population	82.02
A12	Management, bumdes, community, government, financial, system, resources, approach, projects, technology, support, planning, sustainable, challenges, information, local, enterprise, implementation, policy, programme	115.14
A13	Agricultural, food, farmers, agriculture, farm, forest, land, farming, production, climate, change, sustainable, farms, crops, environmental, practices, species, crop, natural, ecological	58.57
A14	Policy, policies, public, management, future, social, resources, rights, impacts, benefits, including, political, India, infrastructure, current, national, resource, world, global, critical	71.79

cultural and social values should be integrated into the planning. The term indigenous is closely related to forest conservation, whereby there are contexts of policy and deforestation, and sustainability (Table 3). This would reflect that deforestation has an impact on indigenous people, whereby policy is needed for their sustainability.

The knowledge structure based on the topic modeling includes term indigenous in the topic of conservation and forest using a topic model with $k = 60$ (C28), as shown in Table 3, while the term forest itself is also embedded in the topic on “environment” and “climate change”.

Table 2 Topic model of village enterprise publication with topic size $k=30$

Topic no	Top terms	Topic intensity
B00	Housing, service, care, health, integrated, retirement, community, services, villages, people, provision, living, sector, public, Australia, private, industry, forms, policy, Australian	15.32
B01	Policy, issues, political, enterprise, policies, recent, implications, differences, historical, period, author, discussed, labour, features, considers, consequences, specific, future, lack, concerned	36.99
B02	Local, tves, ownership, government, township, market, reform, private, property, chinese, governments, rights, privatization, economy, collective, role, reforms, political, transition, sector	97.04
B03	Biomass, technologies, renewable, technology, countries, projects, sustainable, potential, fuel, consumption, modern, policy, electricity, solar, financial, sources, available, provide, Africa, developing	55.48
B04	Costs, governance, organization, banks, integration, trust, actors, recent, institutions, commercial, role, cooperative, forms, learning, common, social, transaction, organizational, competition, effective	26.49
B05	Food, local, management, security, India, resources, organizations, population, Asia, consumption, challenges, increase, foods, natural, human, improved, region, major, people, increasing	29.66
B06	Regional, growth, inequality, province, total, provinces, income, investment, Jiangsu, urban, level, region, regions, increased, developed, coastal, migration, mainly, consumption, urbanization	52.27
B07	Growth, economy, industry, changes, policies, chinese, agricultural, policy, reform, system, structure, industrial, rapid, transformation, countries, world, investment, sector, reforms, agriculture	87.72
B08	Policy, knowledge, including, production, scientific, public, farm, small-scale, ford, regions, poultry, programme, farming, western, England, developed, scheme, changes, images, farmers	25.12
B09	Sector, local, role, business, employment, businesses, policy, economy, growth, activities, opportunities, private, support, policies, services, India, identified, importance, terms, range	79.70
B10	Agricultural, farmers, agriculture, farm, land, production, farming, social, farms, policy, factors, diversification, peasants, farmer, life, security, workers, society, villagers, include5	41.95
B11	Government, housing, urban, family, planning, women, jobs, million, township, capital, local, workers, residents, urbanisation, 1980s, conditions, market, homes, especially, provide	36.83
B12	Social, networks, network, capital, local, process, digital, internet, knowledge, access, mobile, relationships, resources, communities, population, operation, central, innovation, transformation, regional	29.33
B13	Sustainable, forest, sustainability, landscape, ecological, natural, tourism, social, environmental, planning, conservation, forests, species, cultural, resources, cover, land, private, alternative, biodiversity	35.77
B14	Production, market, supply, products, industry, chain, trade, value, marketing, markets, food, agricultural, sector, farmers, export, prices, producers, price, sectors, agriculture	39.71
B15	Efficiency, tves, growth, labor, production, chinese, productivity, performance, industrial, sector, capital, effects, positive, township, period, model, effect, analysis, industry, panel	68.27
B16	Model, system, analysis, management, systems, models, information, application, performance, production, level, variables, evaluation, cost, decision, technology, method, evaluate, agricultural, approach	52.96
B17	Electricity, system, supply, electrification, demand, systems, impacts, grid, load, access, hybrid, benefits, solar, generation, households, cost, business, communities, approach, services	27.18
B18	Innovation, firms, analysis, smes, business, performance, market, environmental, strategic, environment, mechanism, strategies, firm, framework, resources, factors, findings, external, innovations, strategy	45.48
B19	Urbanization, urban, spatial, industrial, system, chinese, cities, process, towns, town, distribution, regional, pattern, factors, villages, city, model, central, culture, tves	46.73

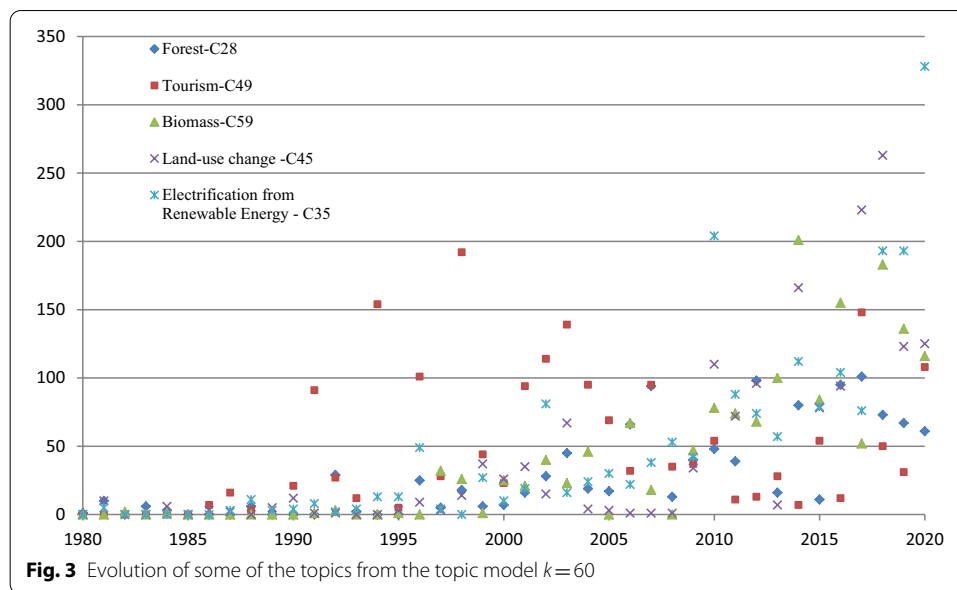
Table 2 (continued)

Topic no	Top terms	Topic intensity
B20	Credit, financial, institutions, access, bank, poor, finance, enterprise, services, loans, market, support, microfinance, financing, lending, women, public, informal, policy, social	39.47
B21	Land, urban, spatial, cultivated, change, farmland, transition, changes, construction, settlements, protection, structure, scale, landscape, land-use, pattern, expansion, conversion, city, rapid	23.59
B22	Production, biogas, khadi, industries, India, crop, plant, rice, total, farmers, yield, cattle, levels, increased, rate, grain, soil, analysis, kvic, respectively	51.46
B23	Environmental, pollution, water, carbon, river, industrial, environment, coal, emissions, impacts, health, protection, township, measures, control, factors, aquifers, safety, national, mines	35.82
B24	Entrepreneurship, business, support, tourism, enterprise, findings, entrepreneurial, entrepreneurs, activity, purpose, implications, factors, owners, literature, qualitative, training, sample, interviews, related, characteristics	58.10
B25	Change, climate, policy, resilience, policies, countries, management, framework, challenges, literature, support, environmental, systems, issues, changes, approach, practices, initiatives, international, strategies	61.84
B26	Income, poverty, women, households, household, activities, livelihood, impact, gender, agricultural, employment, education, nonfarm, labor, farmers, significant, poor, agriculture, migration, level	72.35
B27	Practices, enterprise, understanding, social, framework, analysis, role, processes, approach, theory, empirical, literature, provide, theoretical, practice, emerging, urban, terms, context, particular	81.27
B28	Bumdes, management, community, government, local, financial, implementation, enterprise, villages, accountability, people, district, stakeholders, role, activities, public, communities, resources, potential, governance	41.56
B29	Firms, enterprise, urban, workers, factors, returns, wage, growth, types, Russia, differences, found, firm, experience, performance, market, management, capital, industries, Europe	41.58

Table 3 20 top-terms from some of the topics from the topic model with topic size $k = 60$

Topic no	Top terms	Topic Intensity	Coherence
C28	Forest, species, land, conservation, forests, change, cover, biodiversity, natural, indigenous, populations, diversity, enterprise, forestry, owners, private, nature, deforestation, explored, native	31.675	0.986
C35	Electricity, electrification, solar, renewable, access, demand, benefits, supply, community, India, generation, biomass, communities, households, micro-enterprises, cost, approach, consumption, grid, mini-grids	20.856	0.912
C45	Land, industrial, urban, spatial, cultivated, transition, city, construction, pattern, structure, changes, types, industrialization, land-use, change, patterns, towns, Beijing, built-up, residential	18.794	- 0.986
C49	Tourism, life, residents, examines, living, homes, farmers, location, form, hosts, typical, opportunities, partnership, demand, improving, tourists, understanding, scale, space, home	15.447	- 1.295
C59	Biomass, climate, change, carbon, environmental, sustainable, fuel, emissions, crop, ecosystem, sources, production, resources, global, changes, vulnerability, renewable, wood, increase, crops	17.543	1.808
C26	Innovation, business, firms, role, networks, institutional, products, support, innovations, product, marketing, businesses, external, performance, trust, jatropa, technology, orientation, factors, findings	17.034	- 0.599

The dynamic of the topic models is illustrated using topic model with size $k = 60$, as shown in Fig. 3, whereby the intensity for each topic is calculated using the terms' frequency. Topics of forest, tourism, biomass, land-use change, and electrification



from renewable energy are compared with list of the top-terms are given in Table 3. Most of the topics selected has coherence higher than the mean value of the topic model's coherence (-0.542). Two selected topics are below the mean value, but have higher relevance and easier interpretability. In Fig. 3, the topic of tourism (C49) has shown significant intensities in the 90 s, but since 2003, it has continued to experience a decrease in intensity. On the other hand, the topic of electrification based on renewable energy (C35), especially solar and biomass, continues to develop. The topic of forest (C28) contains aspects of conservation, biodiversity and the utilized species, indigenous, and population, forestry enterprise, private ownership and deforestation. By investigating some selected full-text papers, a relationship can be identified between forestry enterprise and indigenous innovation, namely, through innovation in developing viable businesses for non-timber forest products (de Beer & McDermott, 1989; Matias et al., 2018; Meinhold & Darr, 2019). The term "innovation" in the topic model is embedded in the topic of business innovation (C26), wherein the keywords are such as network, product marketing and technology. These keywords reflect the research goals in investigating a sustainable indigenous rural enterprise through development of network, institutional arrangement, product marketing and the use of technology.

This finding reflects that there are responses from the indigenous people for their resilient livelihood against their vulnerability related to climate change by preserving natural resources and landscape (environment) and forest conservation. The development of tourism in forestry areas is closely related to indigenous people's efforts in supporting forest conservation against land conversion and maintaining biodiversity in the forest, especially the utilized species. In addition, by preserving indigenous traditional food production, they can set up a sustainable market for local products and diversifications. This potential emerges as they developed indigenous knowledge in the use and management of such natural resources. Diversification

from farming (agriculture) can lead to indigenous food production, local tourism services, and the local generation of electricity (such as biogas and biomass). Indigenous innovation in the field of traditional food production is a form of innovation to strengthen the food security of the surrounding community by involving most of the community. Innovations in the forestry and land sectors generally aim at preserving nature, conserving and even using forest products by relying on existing local wisdom. When the community is involved in public services, the main goal is not only to get closer to services, but also to improve the welfare of the local community, usually by streamlining costs and also expanding access to services. In the field of empowerment of small community businesses, the area of study revolves around access to capital, networks, increasing skills and business knowledge of the business actors and also the empowerment and sustainability of local businesses. In terms of basic community needs, indigenous innovation is seen in the form of procurement, electricity, water and the need for employment. In addition, the study of women's empowerment in business is an interesting issue besides leadership, succession in family enterprises, household income and employment. The focus of the study that has received quite a lot of attention is also government policies and programs that can be used as leverage in regional development.

The vision of viewing local resources as opportunity for building indigenous village enterprises should be adopted by all residents, businesses and local government, whereby technology supported innovation will be key to the growth and sustainability. The study's results have shown that government involvement is important in encouraging indigenous knowledge and indigenous innovation through related policy and program development, such as setting up a microgrid based on renewable energy. The local community maintains sustainable and profitable operations from day-to-day electricity use to support a village enterprise. To expand the social entrepreneurship, use the surplus to develop local ecotourism to increase people's welfare. In developing a village enterprise, efforts to maintain business and environmental sustainability include land, forest, water, and river, whereby key terms are technology, innovation, and entrepreneurship. All these key terms described above are identified in Fig. 4, whereby a knowledge structure is attempted to be visualized from the extracting abstracts of the collected articles based on results from topic modeling.

The results show that the topic modeling provides the knowledge structure of the content of researches on village enterprises. The results also show various aspects of indigenous village enterprise, whereby the state of the art is on forest conservation and related environment, ecology, biodiversity, and species sustainability because of the deforestations. Policy development is critical, whereby one of the aspects is the ownership issue (see Table 3). The hidden knowledge that this study found is the identification of economic potentials related to natural resource based indigenous innovations in the sector of agriculture, traditional food production, ecotourism, and renewable energy electrification. To this end, research on the village enterprise should consider and focus on these potentials to elevate the rural economy.

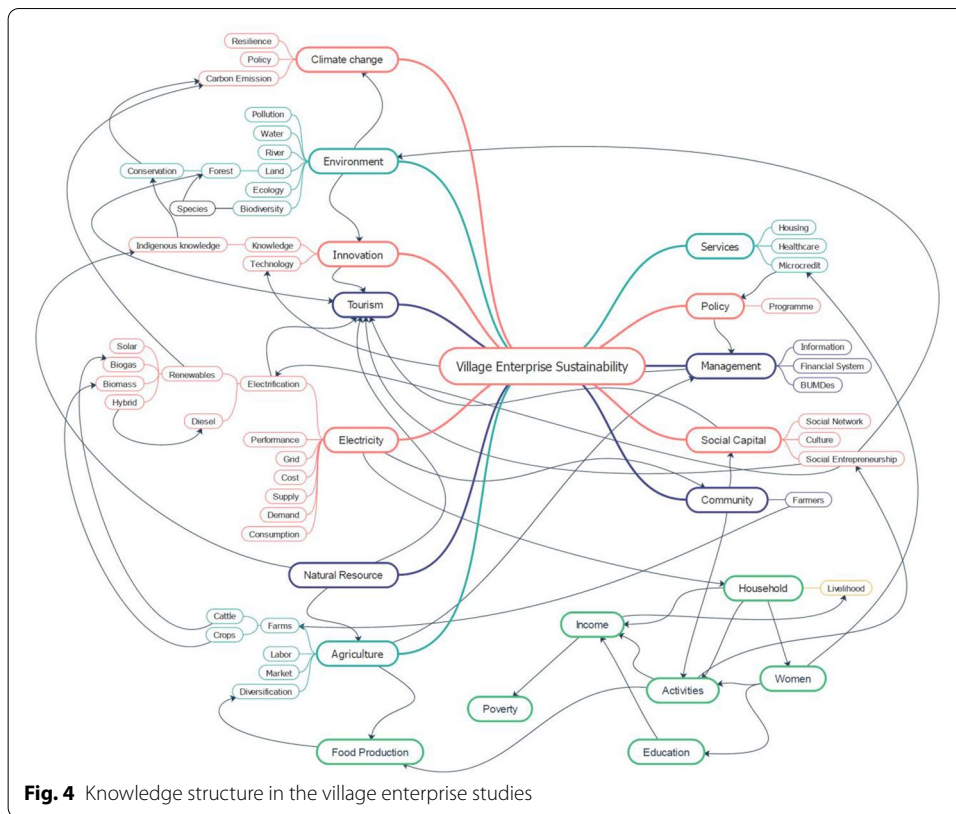


Fig. 4 Knowledge structure in the village enterprise studies

Conclusion

Innovation is a critical factor not only for economic development, social life but also environmental sustainability. The high interest on village enterprise study should be effectively managed by looking at indigenous innovation which is one of the critical factors. There is a gap in the literature that maps the indigenous innovation within the study of the village enterprise. To this end, the authors deploy topic modeling to extract village enterprise research articles from the Scopus and Science Direct repository. The hidden knowledge that this study found is the identification of economic potentials related to natural resource based indigenous innovations in the sector of agriculture, traditional food production, ecotourism, and renewable energy electrification. This finding reflects that there are responses from the indigenous people for their resilient livelihood against their vulnerability related to climate change by preserving natural resources and landscape (environment), and forest conservation. This potential emerges as they developed indigenous knowledge in the use and management of such natural resources. Research on the village enterprise should consider these potentials to elevate the rural economy.

Text analytics using topic modeling application produces knowledge structure of indigenous innovation on the village enterprise through content analysis of research publications from the Scopus and Science Direct repository. This provides a clear and structured picture that empirical studies from indigenous innovation range from traditional food production, local ownership (such as market and property), land, public services (housing, health care, retirement), economy policy, financial micro-credit, environment pollution control, employment, local business sustainability, electricity, women

(gender) and household income, management, and public policy. This study has shown that topic modeling can be used as a tool to the development of a model of knowledge structure in village enterprise which is useful in many areas of study.

Abbreviations

BUMDes: Village owned enterprises in Indonesia; VRIO: Value, Rare, Imitability and Organization analysis framework; LDA: Latent Dirichlet Allocation, a topic modeling model; DTM: Document Term Matrix; TVES: Township and Village Enterprises.

Acknowledgements

Not applicable.

Authors' contributions

RK: conceptualization, methodology, investigation, writing—original draft, project administration; MS: conceptualization, methodology, writing—original draft, software, validation, formal analysis, investigation; AYA: conceptualization, writing—review; RH: data curation, visualization, editing. VJ: conceptualization, review and supervision. All authors read and approved the final manuscript.

Funding

This work was supported by the Ministry of Research and Technology/National Research and Innovation Agency within the scheme Higher Education Excellence Basic Research Grants 2020 under Grant No.: NKB-105/UN2.RST/HKP.05.00/2020.

Availability of data and materials

The data set generated and analysed during the current study are available in the Science Direct and SCOPUS repository.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Faculty of Administrative Science, University of Indonesia, Depok, Indonesia. ²Research Center for STI Policy & Management, Indonesian Institute of Sciences, Jakarta, Indonesia.

Received: 22 January 2021 Accepted: 21 November 2021

Published online: 20 February 2022

References

- Appelbaum, R. P., Gebbie, M. A., Han, X., Stocking, G., & Kay, L. (2016). Will China's Quest for indigenous innovation succeed? Some lessons from nano technology. *Technology in Society*, *46*, 149–163. <https://doi.org/10.1016/j.techsoc.2016.03.004>
- Barney, J. B., & Clark, D. (2007). *Resource based theory. Creating and sustaining competitive advantage*. Oxford University Press.
- Baskaran, S., & Mehta, K. (2016). What is innovation anyway? Youth perspectives from resource-constrained environments. *Technovation*, *52–53*, 4–17. <https://doi.org/10.1016/j.technovation.2016.01.005>
- Blackman, A., & Veit, P. (2018). Titled amazon indigenous communities cut forest carbon emissions. *Ecological Economics*, *153*, 56–67.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022.
- Bonilla, T., & Grimmer, J. (2013). Elevated threat levels and decreased expectations: How democracy handles terrorist threats. *Poetics*, *41*(6), 650–669. <https://doi.org/10.1016/j.poetic.2013.06.003>
- Boon, E. K., & Hens, L. (Eds.). (2007). *Indigenous knowledge systems and sustainable development: Relevance for Africa*. Tribes and Kamla-Raj Enterprises.
- Capel, C. (2014). Mindfulness, indigenous knowledge, indigenous innovations and entrepreneurship. *Journal of Research in Marketing and Entrepreneurship*, *16*(1), 63–83. <https://doi.org/10.1108/JRME-10-2013-0031>
- Chae, B. K., & Park, E. O. (2018). Corporate social responsibility (CSR): A survey of topics and trends using Twitter data and topic modeling. *Sustainability*, *10*(7), 2231. <https://doi.org/10.3390/su10072231>
- Chandra, Y., Jiang, L. C., & Wang, C. J. (2016). Mining social entrepreneurship strategies using topic modeling. *PLoS ONE*, *11*(3), e0151342. <https://doi.org/10.1371/journal.pone.0151342>
- Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Neural information processing systems* 22: 288–296.
- Curry, J. A., Donker, H., & Michel, P. (2016). Social entrepreneurship and indigenous people. *Journal of Co-Operative Organization and Management*, *4*, 108–115. <https://doi.org/10.1016/j.jcom.2016.09.002>
- De Beer, J. H., & McDermott, M. J. (1989). *Economic value of non-timber forest products in southeast Asia*. Netherlands Committee for IUCN.
- Dewalt, B. R. (1994). Using indigenous knowledge to improve agriculture and natural resource management. *Human Organization* *53*(2): 123–31. Retrieved from <http://www.jstor.org/stable/44126875>.

- Ezeanya-Esiobu, C. (2019). Research, innovation, indigenous knowledge and policy action in Africa. In *Indigenous knowledge and education in Africa* (pp. 97–106). Springer Singapore. https://doi.org/10.1007/978-981-13-6635-2_7.
- Gatra, S. (2019). *Total Dana Desa 2019–2024 Rp 400 Triliun*, February 26, 2019. <https://nasional.kompas.com/read/2019/02/26/17333511/total-dana-desa-2019-2024-rp-400-triliun?page=all>.
- Gleb Raygorodetsky. (2018). *Indigenous peoples defend earth's biodiversity—But they're in danger*. November 16, 2018. <https://www.nationalgeographic.com/environment/2018/11/can-indigenous-land-stewardship-protect-biodiversity-/>.
- Howell, A. (2018). Indigenous innovation with heterogeneous risk and new firm survival in a transitioning Chinese economy. *Research Policy*, 44, 1866–1876. <https://doi.org/10.1016/j.rser.2017.05.266>
- Huang, J., Hao, Y., & Lei, H. (2018). Indigenous versus foreign innovation and energy intensity in China. *Renewable and Sustainable Energy Reviews*, 81, 1721–1729. <https://doi.org/10.1016/j.rser.2017.05.266>
- ILO. (2017). *Indigenous peoples and climate change: From victims to change agents through decent work*. International Labour Office, Gender, Equality and Diversity Branch. http://tyndall.ac.uk/sites/default/files/IndigenousPeoplesandClimateChange_0.pdf.
- Jauhiainen, J. S., & Hooli, L. (2017). Indigenous knowledge and developing countries' innovation systems: The case of Namibia. *IJIS*, 1(1), 89–106. <https://doi.org/10.3724/SPJ.1440.101007>
- Karanasios, K., & Parker, P. (2018). Tracking the transition to renewable electricity in remote indigenous communities in Canada. *Energy Policy*, 118, 169–181.
- Li-Ying, J., & Wang, Y. (2015). Find them home or abroad? The relative contribution of international technology in-licensing to 'indigenous innovation' in China. *Long Range Planning*, 48, 123–134. <https://doi.org/10.1016/j.worlddev.2010.05.009>
- Kao, S. W., & Luarn, P. (2020). Topic modeling analysis of social enterprises: Twitter evidence. *Sustainability*, 12(8), 3419. <https://doi.org/10.3390/su12083419>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., et al. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118. <https://doi.org/10.1080/19312458.2018.1430754>
- Makondo, C. C., & Thomas, D. S. G. (2018). Climate change adaptation: linking indigenous knowledge with western science for effective adaptation. *Environmental Science and Policy*, 88, 83–91.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511809071>
- Matias, D. M. S., Tambo, J. A., Stellmacher, T., Borgemeister, C., & von Wehrden, H. (2018). Commercializing traditional non-timber forest products: An integrated value chain analysis of honey from giant honey bees in Palawan, Philippines. *Forest Policy and Economics*, 97, 223–231. <https://doi.org/10.1016/j.forpol.2018.10.009>
- Meinhold, K., & Darr, D. (2019). The processing of non-timber forest products through small and medium enterprises—A review of enabling and constraining factors. *Forests*, 10(11), 1026. <https://doi.org/10.3390/f10111026>
- Mika, J. P., Warren, L., Foley, D., & Palmer, F. R. (2017). Perspectives on indigenous entrepreneurship, innovation and enterprise. *Journal of Management & Organization*, 23(6), 767–773.
- Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 conference on empirical methods in natural language processing*
- Nelson-Barber, S., & Johnson, Z. (2016). Acknowledging the perils of 'Best Practices' in an indigenous community. *Contemporary Educational Psychology*, 47, 44–50. <https://doi.org/10.1016/j.cedpsych.2016.04.001>
- Newman, S.K., & Cavedon, L. (2009). External evaluation of topic models. In *Proceedings of ADCS 2009*, 11–18, Sydney, Australia.
- NgAsike, J. T., & Swadener B. B. (2015). Turkana indigenous knowledge: Environmental sustainability and pastoralist lifestyle for economic survival. In *Indigenous innovation*, 107–27. Leiden, The Netherlands: BrillSense. https://doi.org/10.1007/9789463002264_008.
- Padilla-Meléndez, A., & Ciruela-Lorenzo, A. M. (2018). female indigenous entrepreneurs, culture, and social capital. The case of the Quechua Community of Tiquipaya (Bolivia). *Women's Studies International Forum*, 69, 159–170. <https://doi.org/10.1016/j.wsif.2018.05.012>
- Porter, M. F. (2006). An algorithm for suffix stripping. *Program Electronic Library and Information Systems*, 14(July), 130–137. <https://doi.org/10.1108/00330330610681286>
- Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 248–256).
- Undang-Undang Republik Indonesia Nomor 6 Tahun 2014 Tentang Desa*. 2014. Jakarta.
- Wallach, H. M., Mimno, D. M., McCallum, A. (2009). Rethinking LDA: Why priors matter. In *Advances in neural information processing systems*, 1973–1981.
- Yang, Y., Cai, W., & Wang, C. (2014). Industrial CO₂ intensity, indigenous innovation and R&D spillovers in China's provinces. *Applied Energy*, 131, 117–127. <https://doi.org/10.1016/j.apenergy.2014.06.033>
- Zhao, S. L., Cacciolatti, L., Lee, S. H., & Song, W. (2015). Regional collaborations and indigenous innovation capabilities in China: A multivariate method for the analysis of regional innovation systems. *Technological Forecasting and Social Change*, 94, 202–220. <https://doi.org/10.1016/j.techfore.2014.09.014>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.