

LDA-Based Topic Mining for Unveiling the Outstanding Universal Value of Solo Keroncong Music as an Intangible Cultural Heritage of UNESCO

Denik Iswardani Witarti^{1*}, Danis Sugiyanto², Atik Ariesta³, Pipin Farida Ariyani⁴, Rusdah⁵

Faculty of Communication and Creative Design, Universitas Budi Luhur, Jakarta, Indonesia¹

Faculty of Performing Art, Institut Seni Indonesia Surakarta, Surakarta, Indonesia²

Faculty of Information Technology, Universitas Budi Luhur, Jakarta, Indonesia^{3,4,5}

Abstract—Outstanding Universal Value (OUV) is an essential value of culture and nature. It is so extraordinary that it transcends national boundaries and becomes generally crucial for all humanity's current and future generations. A culture with this value needs permanent protection because it is considered a critical heritage for the world community. Solo keroncong music, as one of the local wisdom owned by the Indonesian nation, has yet to be recognized as one of the UNESCO Intangible Cultural Heritage (ICH). It has even become an instrument of Indonesia's soft power diplomacy in several countries, such as Malaysia, England, and the United States. It must be of OUV and meet at least one of ten selection criteria to be included on the World Heritage List. This study explored the OUV of Solo keroncong music using Latent Dirichlet Allocation. The primary data were obtained by conducting an FGD with the Indonesian Keroncong Music Artist Community (KAMKI) Surakarta and in-depth interviews with several keroncong figures in Solo. The result showed there are four topics with a coherent score of 0.51. Then, the expert mapped those four topics into three OUVs of Solo keroncong music as temporary findings. Keroncong music is a masterpiece of human creativity, a witness to civilization, and has traditional values. These findings showed that Solo keroncong music is worthy of being proposed as one of the UNESCO ICH.

Keywords—LDA; OUV; Solo keroncong; text mining; topic modeling

I. INTRODUCTION

The United Nations Educational, Scientific and Cultural Organization (UNESCO) is a unique agency in the United Nations (UN) that deals with education, science and culture. UNESCO programs are divided into five main sectors: Education, Natural Sciences, Social Sciences, Culture, and Communication and Information [1]. Indonesia ratified the Convention for the Safeguarding of Intangible Cultural Heritage issued by UNESCO in 2003, which was then ratified in Presidential Regulation 78 of 2007. Consequently, Indonesia must protect its cultural wealth. Indonesia is required to record cultural works as a protection effort. The Directorate of Cultural Heritage and Diplomacy will determine and then grant status as Warisan Budaya Tak Benda (WBTB) or Intangible Cultural Heritage (ICH) by the minister based on the recommendations of the team of experts formed. According to the 2003 UNESCO Convention, what is included in Intangible Cultural Works are Traditions and oral expressions, including language as a vehicle for intangible cultural heritage;

Performing arts; Customs, rituals, and celebrations; Knowledge and behavioral habits regarding nature and the universe; Traditional craft skills [2], [3]. Music and song are present in 304 of the 584 elements on the list (52%), referring to music alone or combined with other dimensions such as dance and poetry [2].

Keroncong music is one of Indonesia's cultural arts riches that still exists in the era of the development of the modern music industry. During the Dutch era, Keroncong Tugu, the origin of Indonesian keroncong music, was famous for entertainment in Batavia. Keroncong music spread widely and blended with gamelan, the primary musical culture in Java at that time. Combining keroncong musical instruments with the musicality of gamelan later became the characteristic of Solo keroncong music. The music played is more relaxed (nglaras in Javanese) than Keroncong Tugu, which developed on the coast. The Indonesian Keroncong Music Artist Community (KAMKI) of the Surakarta City Branch Representative Council (DPC Surakarta) noted that 60 community members were actively preserving keroncong music in Solo City [4]. The strains of Solo keroncong have also succeeded in attracting the interest of music lovers from abroad. Unfortunately, the beauty of the Solo keroncong music has not been included in the UNESCO ICH list. UNESCO registration will also impact Indonesia's positive image as a nation preserving cultural wealth. International recognition of keroncong music will also increase the appeal of cultural tourism in Solo. Foreign tourists will be interested in coming to Indonesia and directly enjoying the musical strains of Solo keroncong music and other local cultures.

Based on the background above, this study was designed to explore the feasibility of keroncong music as one of the ICH by exploring the Outstanding Universal Value (OUV) contained in Solo keroncong music. OUV is an essential value of a culture or nature that is so extraordinary that it transcends national boundaries and becomes generally crucial for all humanity's current and future generations.

Exploring the OUV of Solo Keroncong Music used ten criteria based on the Operational Guidelines for implementing the World Heritage Convention (OGIWHC) [1]. The research involved KAMKI Surakarta in gaining in-depth knowledge about Solo Keroncong Music. The results of discussions and interviews were processed into texts for use in LDA modeling.

The modeling results in several topics about the main values of Solo keroncong music. These main topics' values were evaluated based on OGIWHC to see whether Solo keroncong music meets the OUV criteria as a requirement for UNESCO ICH determination.

This study is organized into five sections. The first section briefly describes Keroncong music, the background to the problem area, the objective of the study, and the significance of the research output. The second section deals with previous studies on soft power diplomacy and topic modeling using LDA in terms of ICH. The third section discusses the methodologies used to conduct this study. The fourth section is about the analysis and results of this study. Finally, the last section discusses the findings and recommendations for future work.

II. RELATED WORK

A. Soft Power Diplomacy

This study is anchored in the concept of soft power diplomacy, a framework developed by Joseph Nye. It posits that culture, politics, and foreign policy are the three pivotal sources of soft power. Effectively utilizing these sources is instrumental in determining diplomacy's success [5]. Soft power diplomacy is a strategy to influence foreign public opinion and behavior by utilizing the appeal of culture, values, and inclusive foreign policies. The implementation of soft power diplomacy pursues the goal of creating a positive image of the country and expanding its influence in the international world. A country's reputation and how the target public accepts it will determine the success of soft power diplomacy. The appeal and influence result from a social process involving both parties, so the effects of soft power will only occur if there is a mutually impactful relationship between the two parties [6].

Researchers have recently conducted studies using soft power to implement Indonesian diplomacy. Research on the peaceful message of the XX Papua National Sports Week (PON) was carried out to highlight sports as one of the instruments of soft power diplomacy [7]. The study results show the success of the government's soft power diplomacy in warding off security issues in Papua through the success of PON XX. The researcher also conducted a sentiment analysis of the news of the K-Pop concert on the 50th Anniversary of bilateral relations between Indonesia and South Korea. The positive sentiment of the Indonesian public explains that music can be an instrument of soft power diplomacy in harmonizing bilateral relations between Indonesia and South Korea [8]. Currently, the researcher is conducting research (ongoing) on the potential of keroncong music as cultural diplomacy between Indonesia and Malaysia [9]. Initial results found the public's love abroad, especially in Malaysia, for the beauty of the Solo keroncong music [9]. This is the background for the researcher to develop in a new study to find the primary value of Solo keroncong music. The researcher has conducted a study on the evolution of keroncong music in Richmond, Virginia. The Rumpu Orchestra is an example of the rapid development of keroncong music outside Indonesia. This orchestra creates new music by combining various Keroncong music cultures, such as Javanese, Balinese, Sundanese,

Appalachian, and Irish [10]. The research also involved the keroncong community in Solo City as one of the determinants of the success of Indonesia's soft power diplomacy. The involvement of non-state actors is an effective tool for implementing soft power more efficiently [11].

B. Topic Modeling using LDA in Terms of ICH

Reference [12] stated that the development of new media has enabled intangible cultural heritage to be disseminated through online platforms and attracted the attention of many contemporary young people. Classification and discussion on the value of intangible cultural heritage were essential ways to help with inheritance and dissemination. Real online reviews were collected using the Bilibili website as the research data source. A text-based BiGRU-Attention model was conducted to achieve value recognition and classification, and keyword statistics and topic analysis were performed for a score of more than 77%. The Cultural Value Perception (CVP) category has the best classification performance. Through the topic analysis of comments and keywords, the cultural value of intangible cultural heritage is its core connotation, social value is the primary purpose, and economic value is the power source [12].

Meanwhile, based on cultural identity and sovereignty, intangible cultural heritages (ICHs) are disappearing at an alarming rate and facing an existential crisis [13]. With the emergence of neural network models, the development of digital technology has revived many things that were on the verge of extinction. Traditional cultures and industries that initially seemed unrelated can take on new forms with the help of digital technology, thus enabling ICHs to find new ideas for development. The study took Huizhou ICH as an example and tried to design and construct a Huizhou ICH database and a digital map of Huizhou ICH, establishing a database for management and operation. The paper applies information space theory to study the use of ICH's digital resources under the threshold of the neural network. It employed digital information technology to recode, reconstruct, and interpret ICH. As a result, traditional ICH items were displayed digitally, improving the public's recognition of digital ICH items, thereby promoting the inheritance and dissemination of ICH [13].

The study in [14] designed and constructed a knowledge ontology framework for sports intangible cultural heritage (ICH) resources to support their preservation and inheritance by integrating and mining sports ICH resources. The study collected multiple data on sports ICH from various data sources and constructed ICH knowledge ontology using the CIDOC CRM metadata reference model and seven-step method. To enrich the content of the ontology, the TextRank algorithm was used to extract critical textual information and design a domain-specific NER model for sports NRM. In addition, the study adopted the VSM vector space model for text representation and used a proven hierarchical classification model for text categorization to improve classification accuracy. The study also explores the similarity calculation of concepts in the sports NRM ontology and proposes a semantic similarity calculation formula based on the ontology concepts. Respondents' willingness to pay was investigated through the conditional value method (CVM) to assess the value of sports NRM tourism resources. Finally, the factors influencing

respondents' willingness to pay were analyzed using statistical analysis and a logistic regression model, and it was found that they were mainly influenced by the degree of understanding of sports non-heritage resources and the level of education. The results of this study not only provide theoretical and methodological support for the effective integration and excavation of sports non-heritage resources and a new perspective on their protection, inheritance, and sustainable development [14].

ICH has important historical, cultural, and spiritual values. The study in [15] used the intangible cultural heritage of Yunnan Painting, the national architectural decoration of Dali, as the case study object. They crawled the consumer's comment text data to establish the Yunnan Painting consumer comment text dataset. Secondly, using the K-means method, they clustered high-frequency words from the dataset using text mining technology, TF-IDF keyword extraction, and LDA themes. Finally, they evaluated the economic potential of Yunnan Painting in terms of consumers' cognitive abilities and the emotional appeal of its products based on the extracted results. The study revealed that the dressmaking, painting, and tourism industries were the main economic activities of Yunnan color painting, with the tourism industry requiring further strengthening. According to the text mining of high-frequency words, the frequency of "product" in Yunnan painting was the highest at 296 times, followed by "culture," "technology," "cultural creativity," and "theme." Analyzing consumers' emotional text features reveals that most consumers have positive emotions towards Yunnan-colored paintings, while very few have negative emotions. The proposed economic optimization strategy is based on the results and provides a reference value for the economic development of other intangible cultural heritage [15].

The study in [16] conducted a paper's topic-tracking research based on the topic model. Firstly, the LDA model was used to extract the topic information from the news texts of different time windows. Then, the improved Single-Pass algorithm was used for topic tracking, in which the time decay function and the JS divergence were used to measure the similarity between the topics. Finally, the content and strength of the topics were analyzed for the results of topic tracking. They found that the topics discovered by the LDA model were more reliable than the k-means clustering in topic recognition. For topic tracking, the perplexity degree determined the optimal number of topics in the time window [16].

The combination of machine learning-driven topic modeling using Latent Dirichlet Allocation (LDA) and network analysis techniques examined a corpus of Korean and Japanese research papers on ICH [17]. LDA topic modeling identified three primary themes: technology and ICH, safeguarding ICH, and methodologies and approaches in ICH research.

Previous studies have been conducted related to analyzing [12] and exploring the value of ICH [14], [15] and its digitalization [13]. Unfortunately, no one discussed topic modeling to find UNESCO's Outstanding Universal Values of those ICHs. Topic detection or topic modeling has been conducted using LDA [16] and machine learning [17]. In terms of the dataset, this study used the primary data as the result of

the FGD process, while the previous study used secondary data by crawling [15].

III. METHODOLOGY

This study explored the OUV of Solo keroncong music, which has not been explored before. Exploratory research helps develop a phenomenon that is not widely known or has little information [18]. Exploratory methods allow researchers to understand complex social phenomena more deeply before formulating specific research questions [19]. Fig. 1 shows the proposed methodology used in this study.

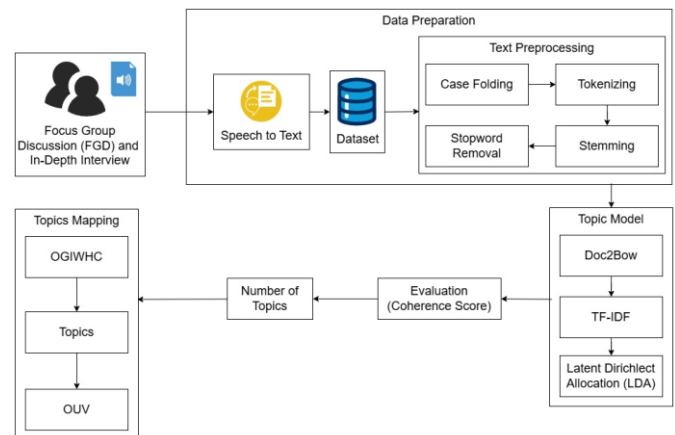


Fig. 1. Proposed methodology.

A. Data Collecting

The first step is to collect primary data. The study used primary data from conducting Focus Group Discussion (FGD) with the Surakarta Keroncong Music Artist Community (KAMKI) and in-depth interviews with several keroncong music figures and activists in Solo. FGD is a structured group discussion with various perspectives or experiences related to a particular research topic [20]. In-depth interviews are qualitative data collection methods that allow researchers to understand individuals' experiences, views, and perspectives related to the research topic [21]. FGD involved 40 people, and in-depth interviews involved five informants. They discussed the main values of keroncong music so that it can be recognized as UNESCO ICH. Initial data were obtained through audio recordings of FGD and in-depth interviews. There were six files of audio recordings.

B. Data Preparation

The recording results were then converted into text using TurboScribe.ai to create a dataset. The text obtained from TurboScribe.ai was then corrected for formats such as spaces, new lines, and inappropriate words. The informants verified the transcript to ensure the validity of the FGD and in-depth interview results. After the transcript (dataset) is verified, the following step is to prepare the data. The dataset contains eight documents. Text processing consists of case folding, tokenizing, stemming, and stopword removal. Case folding is the activity of removing special characters and changing them to lowercase. Tokenizing breaks down text based on spaces into a word (token). Stemming maps each token into a basic word form for nouns, verbs, and adjectives [22]. Stopword

removal is the activity of removing words that refer to words in a language (this study uses the Indonesian language) that are grammatically irrelevant to the content of the text, such as "di", "yang", "dan", "maka", and others [22]. The Sastrawi Python library implements the Nazief and Adriani algorithms used during the stemming and stopwords removal process because it provides good results for the stemming process of Indonesian-language documents [23]. The results of this phase are transcripts datasets of the text on the FGD activities and in-depth interviews.

C. Modeling

Modeling begins by determining the corpus (collection of documents) using the Gensim Doc2Bow library. Then, text weighting and topic modeling follow.

Term (or text) weighting refers to calculating and determining weights. Term Frequency-Inverse Document Frequency (TF-IDF) is the most popular scheme in text weighting. The weight of words calculated by the TF-IDF scheme is proportional to the frequency of their occurrence in a particular text but inversely proportional to the frequency in other texts [22]. TF-IDF is an unsupervised statistical algorithm commonly used in information retrieval and text mining to assess the importance of words in a particular document in a corpus [15].

TF, or word frequency, represents the number of times a word appears in a given document, and this number is usually normalized to avoid TF bias towards long documents. Defining any word as T_i , the TF of the word T_i is formulated as [15]:

$$TF_{i,j} = \frac{N_{i,j}}{\sum_k N_{k,j}} \quad (1)$$

$N_{i,j}$ is the number of times the word T_i appears in a document D_j , and $\sum_k N_{k,j}$ is the total number of times all words appear in a document D_j .

IDF is the inverse document frequency, representing the distribution of documents containing a particular word in the corpus. The IDF of the word is formulated as [15]:

$$IDF_i = \log \frac{|D|}{1+|\{j:T_i \in D_j\}|} \quad (2)$$

$|D|$ is the total number of documents in the corpus, $1+|\{j:T_i \in D_j\}|$ is the number of documents containing the word T_i 1 is added to avoid the denominator being 0 (i.e., all documents do not contain the word).

A statistical-based technique is used to assess the importance of a word in a document to the corpus, namely Term Frequency-Inverse Document Frequency (TF-IDF). This technique calculates the weight of a word by measuring the frequency of its appearance in the document and its relevance to all documents in the corpus [22]. The calculation of TF-IDF is shown in (3). The larger the TF, the smaller the IDF is. In other words, the more times a word appears in a document, the fewer times the word appears in the corpus. The higher the value of TF-IDF, the more important the word is to this article and the more it can represent the article. TF-IDF is calculated as [15]:

$$TFIDF_{i,j} = TF_{i,j} \times IDF_i = \frac{N_{i,j}}{\sum_k N_{k,j}} \times \log \frac{|D|}{1+|\{j:T_i \in D_j\}|} \quad (3)$$

The results of TF-IDF weighting are in the form of a term-document matrix (TDM) and a list of essential words with the highest weight in each document. TF-IDF weighting is based on the corpus.

The next step is topic modeling. Topic modeling is one of the most widely used techniques in the unsupervised machine-learning field [24], [25], [26], [27]. According to [24], topic modeling assumes that a set of hidden topics structures each corpus document. Applying this method, researchers can uncover latent topics and find relationships among topics from textual data. Many social scientists have used topic modeling, especially Latent Dirichlet Allocation (LDA), the most popular method for topic modeling [27], [28], [29].

LDA is a generative probabilistic model of corpus. The LDA method, first introduced in [24], estimates the probability of topic distribution in documents and word distribution in topics [25]. LDA requires an input parameter (k) to determine the number of topics generated. For the LDA model, the number of topics (k) should be determined by researchers [24], [29]. This study determines the number of topics (k topics) by looking at the coherence score [26]. The coherence score indicates semantic similarity between words on a topic [30].

LDA is a method that assesses that each document is represented as a probability distribution over latent topics, and distribution over words characterizes each topic [24]. Words with the highest probability in each topic are usually used to determine its topic, which, in this case, are the main values. LDA can easily assign probability to a new document; no heuristics are needed for a new document to be endowed with a different set of topic proportions than those associated with the training corpus documents. Based on the cooccurrence of words in a large corpus, LDA algorithms simultaneously estimate topics and assign topic weights to each document. The outputs from a topic modeling algorithm are lists of weighted words, where each list is a topic and higher weighted words in a list are more indicative of that topic. It represents each document as a distribution over topics, which can be used to detect semantic patterns across all documents [24].

In Fig. 2, α and β are determined empirically. Meanwhile, z and θ are determined by the LDA model. M is the number of documents; N is the number of feature words in document m . α represents the priori parameter of the topic distribution over the entire document set. β represents the priori parameter of the word distribution across all topics.

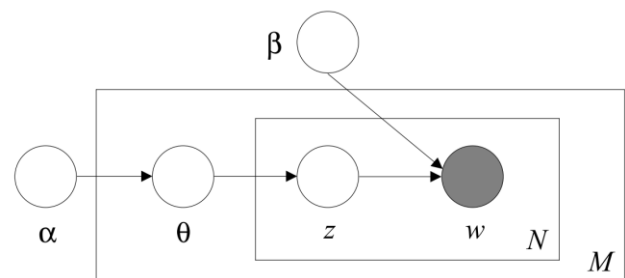


Fig. 2. LDA model diagram [24].

Based on α , the distribution θ of topics in the document is generated. θ is the polynomial distribution of the topic in document M . A topic z is selected from the distribution θ in a document. z is the topic of the n^{th} word in document m ; w is the n^{th} word in document m [16]. Algorithm 1 shows the LDA generation process.

Algorithm 1: LDA Generating Model

```
1: for all topics  $k \in [1, K]$  do
2:   sample mixture components  $\phi_k \sim \text{Dir}(\beta)$ 
3: end for
4: for all documents  $m \in [1, M]$  do
5:   sample mixture proportion  $\theta_m \sim \text{Dir}(\alpha)$ 
6:   sample document length  $N_m \sim \text{Poiss}(\epsilon)$ 
7:   for all words  $n \in [1, N_m]$  do
8:     sample topic index  $z_{m,n} \sim \text{Mult}(\theta_m)$ 
9:     sample item for word  $w_{m,n} \sim \text{Mult}(\phi_{z_{m,n}})$ 
10:  end for
11: end for
```

D. Evaluation

LDA algorithms assume there are n topics across all documents to identify topics in a set of documents. The distributions of words determine the distributed k topics across all the documents. This study measures the number of topics (k topics) by the Coherence score [26]. Conversely, the coherence score shows the semantic similarity between words on a topic [30]. The coherence score is formulated as [30]:

$$\text{score}(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)} \quad (4)$$

where,

$D(v_i, v_j)$ counts the documents containing the words v_i and v_j .

$D(v_j)$ counts the number of documents containing the word v_j .

E. Topics Mapping

The primary value is identified by examining the probability of weighing words by the number of topics determined. These main topics' values were evaluated based on OGIWHC to see whether Solo keroncong music meets the OUV criteria as a requirement for UNESCO ICH determination.

IV. ANALYSIS AND RESULT

The dataset used in this study consists of eight documents. After going through the text processing, 2,563 words were obtained, with details of 86 words in the first document, 217 in the second document, 322 in the 3rd document, 480 in the 4th document, 304 in the 5th document, 375 in the 6th document, 268 in the 7th document, and 511 in the 8th document. The weight of each word was then calculated using TF-IDF. Table I shows some examples of words with frequency and TF-IDF values.

TABLE I. EXAMPLE OF TERM FREQUENCY AND TF-IDF SCORE

Document No.	Words	Term Frequency	TF-IDF Score
0	<i>baku</i>	9	0.5588422545
0	<i>adaptasi</i>	6	0.372561503
0	<i>patah</i>	4	0.372561503
0	<i>ciri</i>	2	0.1862807515
0	<i>moresko</i>	2	0.1862807515
1	<i>anggota</i>	7	0.3556741263
1	<i>asosiasi</i>	6	0.3048635368
1	<i>sumbang</i>	8	0.2709898105
1	<i>perilaku</i>	5	0.2540529474
1	<i>kamki</i>	13	0.220179221
2	<i>rekam</i>	10	0.261233166
2	<i>abad</i>	8	0.2089865328
2	<i>belanda</i>	5	0.1959248745
2	<i>hindia</i>	5	0.1959248745
2	<i>lokananta</i>	5	0.1959248745
3	<i>naskah</i>	17	0.2863157107
3	<i>kait</i>	14	0.2357894088
3	<i>masuk</i>	27	0.2273683585
3	<i>lobi</i>	8	0.2021052076
3	<i>presentasi</i>	8	0.2021052076
4	<i>karakter</i>	8	0.361648169
4	<i>ahli</i>	10	0.2132273836
4	<i>dinas</i>	6	0.1808240845
4	<i>sederhana</i>	4	0.1808240845
4	<i>wayang</i>	8	0.1705819069
5	<i>cengkok</i>	11	0.3362903237
5	<i>ritmis</i>	9	0.2751466285
5	<i>biola</i>	8	0.2445747809
5	<i>ngeroncong</i>	5	0.152859238
5	<i>pakem</i>	5	0.152859238
5	<i>senggaan</i>	5	0.152859238
6	<i>ajar</i>	24	0.3947967127
6	<i>ramlee</i>	8	0.2790008836
6	<i>Malaysia</i>	15	0.2467479454
6	<i>zaman</i>	7	0.2441257732
6	<i>melayu</i>	13	0.2138482194
7	<i>catat</i>	21	0.2236946025
7	<i>kait</i>	9	0.1917382307
7	<i>seni</i>	22	0.1589039312
7	<i>seniman</i>	10	0.150731548
7	<i>wbtb</i>	12	0.1278254871
7	<i>naskah</i>	6	0.1278254871
7	<i>tempe</i>	6	0.1278254871
7	<i>instrumentasi</i>	4	0.1278254871
7	<i>publik</i>	4	0.1278254871
7	<i>wonten</i>	4	0.1278254871

After the term was weighted, modeling was carried out with the LDA. Several experiments with topics 2 to 10 were carried out to obtain the best Coherence Score. Fig. 3 shows that the number of topics 4 obtains the best Coherence Score of 0.51.

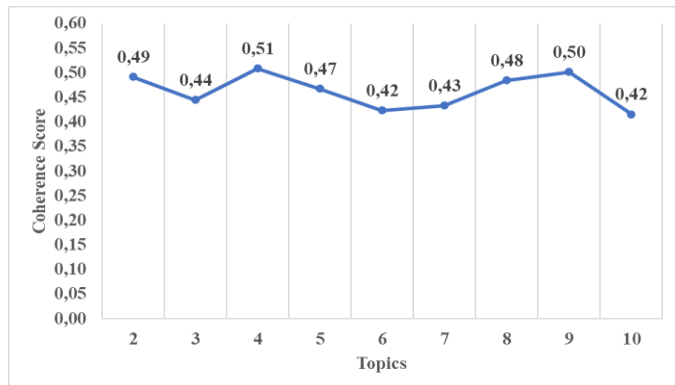


Fig. 3. Comparison of coherence score.

The higher the coherence score, the better the topic modeling interpretation results produced. Using four topics (k=4), the model gave the semantic similarity between words on a topic. Thus, every topic was visualized using pyLDAvis and the word cloud, which was needed in the analysis process.

The terms in the topic modeling are shown in the text, which is primarily frequent in the document. These were depicted by the circle size (as seen in Fig. 5, 7, 9, and 11). Representation of the result using a scatter plot would reveal the distance between topics, the distribution, and the relationship between topic levels. The distance between two or more topics approximates their semantic relationship. Note that close topics 1, 2, 3, and 4 are semantically related, which describes the terms in the topics. As observed in Fig. 5, 7, 9, and 11, the terms are described in the articles concerning the topic's distribution. This reveals that topics 1, 2, 3, and 4 are semantically distributed and have a relationship on topic levels. These reveal four selected topics from the topic model analyzed using the LDA model. The LDA model was one of the input arguments, along with the corpus and dictionary of the emerging terms used for the topic modeling. The slider (λ) in the web-based interactive visualization depicts the relevance metric of the rank terms. It is worth knowing that the terms of the topic are ranked in decreasing order by default following the topic-specific probability ($\lambda = 1$). Fig. 5, 7, 9, and 11 reveal the standard terms from the topic model when the slider is at full probability.

Fig. 4 shows the word cloud generated by topic 1, and Fig. 5 shows the model visualization. The comments containing the keywords "kait" and "naskah" appear the most often. In addition, expressions related to music such as "cengkok", "biola", "ritmis", and "genre" are also frequently mentioned.

The model visualization of topic 1 (Fig. 5) reveals the top 30 most relevant terms for topic 1, highlighted with the red bar. These terms account for 49.6% of the overall term frequency, highlighted with the blue bar. Those terms are kait, naskah, masuk, karakter, catat, wakil, usul, ahli, lobi, presentasi,

Menteri, WBTB, UNESCO, untung, budaya, kategori, seni, mohon, wayang, dinas, program, pihak, waris, sederhana, eksklusif, akademik, nyata, sapto and dulu.



Fig. 4. Word cloud of topic 1.

Fig. 6 shows the word cloud generated by topic 2, and Fig. 7 shows the model visualization. The comments containing the keywords "baku", "adaptasi", "karakter", and "patah" appear the most often. Other expressions related to the community such as "anggota", "asosiasi", "ahli", "KAMKI", and "perilaku" are also frequently mentioned.

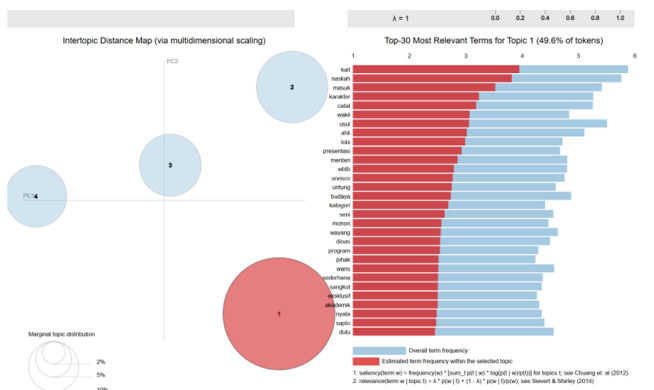


Fig. 5. Model visualization of topic 1.



Fig. 6. Word cloud of topic 2.

The model visualization of topic 2 (Fig. 7) reveals the top 30 most relevant terms for topic 2, highlighted with the red bar. These terms account for 20.2% of the overall term frequency, highlighted with the blue bar. Those terms are baku, adaptasi, patah, usul, KAMKI, cengkok, target, ciri, Moresko, seniman, anggota, gaya, perilaku, asosiasi, sriwedari, sumbang, wadah,

gemar, khas, ndak, ritmis, catat, sumbangsih, wayang, ahli, stimulant, budaya, artis, biola, and enggak.

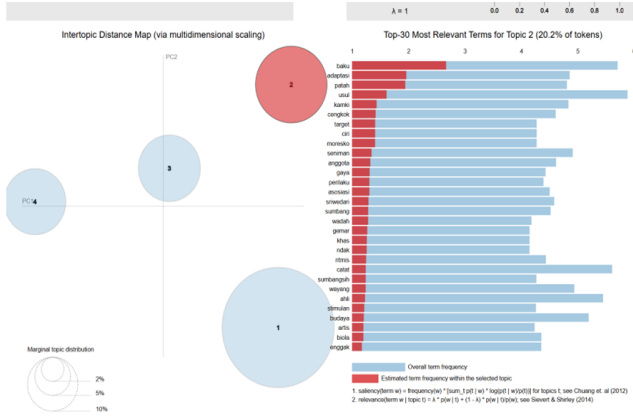


Fig. 7. Model visualization of topic 2.

Fig. 8 shows the word cloud generated by topic 3, and Fig. 9 shows the model visualization. The comments containing the keywords “ajar” and “Ramlee” appear the most often. Other expressions related to history and tradition, such as “Malaysia,” “Melayu,” “zaman,” “kumpul,” and “patriotik,” are also frequently mentioned.



Fig. 8. Word cloud of topic 3.

The model visualization of topic 3 (Fig. 9) reveals the top 30 most relevant terms for topic 3, highlighted with the red bar. These terms account for 15.2% of the overall frequency, highlighted with the blue bar. Those terms are ajar, Malaysia, cengkok, ritmis, lagu, Ramlee, zaman, and biola. Topic 3 mostly contains terms related to music and the instrument such as cengkok, ritmis, lagu, biola, vokal, klasik, gitar, genre, akor, senggaraan, and ngeroncong; the value of education, tradition, and history such as ajar, Malaysia, zaman, Ramlee, Melayu, pakem, baku, patriotik, metode, kumpul, karakter, mula, and pengaruh.

Fig. 10 shows the word cloud generated by topic 4, and Fig. 11 shows the model visualization. The comments containing the keywords “rekam” and “abad” appear the most often. Other expressions related to the history such as “Hindia”, “Belanda”, “lokananta”, and “historis” are also frequently mentioned.

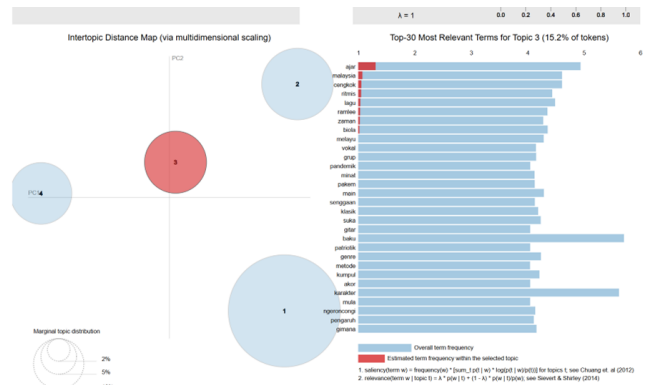


Fig. 9. Model visualization of topic 3.



Fig. 10. Word cloud of topic 4.

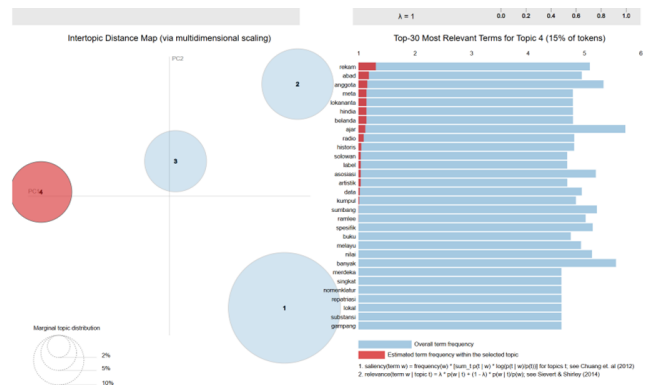


Fig. 11. Model visualization of topic 4.

The model visualization of topic 4 (Fig. 11) reveals the top 30 most relevant terms for topic 4, highlighted with the red bar. These terms account for 15% of the overall frequency, highlighted with the blue bar. Those terms are rekam, abad, anggota, meta, lokananta, Hindia, Belanda, ajar, radio, historis, solowan, label, asosiasi, artistic, data, and kumpul.

LDA modeling results in four topics with the highest coherence score of 0.51. The legitimate source analyzed each word in every topic and mapped it into the OUV values. These main topics' values were evaluated based on OGIWHC to see whether Solo keroncong music meets the OUV criteria as a requirement for UNESCO ICH determination. The result can be seen in Table II.

TABLE II. MAPPING TOPICS TO OUV VALUES

No. OUV	OUV Value	Topic 1	Topic 2	Topic 3	Topic 4
1	Masterpiece	<i>Cengkok Ritmis Biola Genre Senggaan</i>	<i>Karakter Baku Adaptasi Anggota Asosiasi Kamki Ahli Moresko</i>	<i>Ramlee Melayu Minat Grup Vokal Suka Lagu</i>	<i>Rekam Abad Hindia Belanda Lokananta Artistik Radio Historis</i>
3	Civilization	<i>Kait Naskah Ritmis Catat</i>	<i>Adaptasi Sriwedari</i>	<i>Ajar Zaman Melayu Patriotik Grup</i>	<i>Rekam Abad Hindia Belanda Lokananta Radio Historis Buku</i>
6	Tradition	<i>Kait Usul</i>	<i>Adaptasi Wayang Usul Perilaku Moresko</i>	<i>Ajar Ramlee Melayu Patriotik Vokal Lagu Suka</i>	<i>Rekam Hindia Belanda Histori Spesifik Nilai</i>

V. CONCLUSION

This research explored how the data were collected using FGD and in-depth interviews. These primary data are the strength of this study, as Solo keroncong music has never been explored before. Thus, the dataset became this study's key finding and novelty. This research is focusing on creating the dataset and implementing the LDA model. Future research exploring other AI-based cultural studies is still challenging.

The study aimed to find the main values contained in Solo keroncong music as a requirement for eligibility to be proposed as one of UNESCO's intangible cultural heritages. Recognizing Solo keroncong as UNESCO's intangible cultural heritage will have positive implications for preserving Indonesian culture: (1) International recognition will improve Indonesia's image as a country with a rich cultural heritage. (2) Indonesian society, especially the younger generation, will appreciate and be proud of keroncong music as part of their national identity, so the younger generation is more interested in studying and developing keroncong music through innovation and collaboration with other music genres. (3) UNESCO's recognition can increase tourists' interest in learning more about keroncong music, which impacts the creative economy sector. (4) Keroncong music will be better documented, including recordings, academic research, and digital archives.

The corpus was first determined using the Gensim Doc2Bow library. Then, term weighting was conducted by TF-IDF, and topic modeling was used LDA. The modeling phase resulted in four topics, which were then mapped into three OUVs. The results of the research found that there are three OUVs possessed by Solo keroncong music, namely categories (1) representing a masterpiece of human creative genius; (3) bearing a unique or at least exceptional testimony to a cultural

tradition or to a civilization which is living or which has disappeared; (6) being directly or tangibly associated with events or living traditions, with ideas, or with beliefs, with artistic and literary works of outstanding universal significance. (The Committee considers that this criterion should preferably be used in conjunction with other criteria). The following is the explanation of the value of the eligibility of Solo keroncong music as UNESCO's intangible cultural heritage based on the results of this research.

A. Keroncong Music is a Masterpiece of Human Creativity

Solo keroncong music is worthy of being a masterpiece of human creative genius because it combines various cultural elements, musical techniques, and unique artistic values. Keroncong music was born due to a harmonious blend of local Indonesian culture with Portuguese fado music. This music became a cultural heritage developed through acculturation with Javanese, Malay, and other cultures. Keroncong music creatively underwent a process of adaptation to local culture. Despite being influenced by outside influences, keroncong music developed with a local taste in melody, instruments, and lyrics.

The main value of keroncong music lies in its artistic value. Keroncong music has unique characteristics in terms of melodic rhythm. The main instruments commonly used in keroncong music are cak, cuk or ukulele, guitar, cello flute, and violin, creating a distinctive sound and unique musical aesthetics. If keroncong music is suspected of being influenced by the presence of Tugu keroncong music, it turns out that there are abnormalities here and there. Keroncong music, which is currently developing in the mainstream, is not like the music in Tugu. There are differences because the music developed in the archipelago is explicitly treated in each region. For example, what developed in Surakarta already has mainstream gamelan music. In the interior of Java, especially Surakarta, music that developed more slowly when using instruments like those above was appreciated and created in such a way according to the character of Javanese. In various regions in the Indonesian archipelago, keroncong music developed with various variants, some using the chordal system like Western music, some using rhythms or in the local scale area. In the Surakarta distribution area, a type of Javanese Langgam uses the Slendro and Pelog scales. Keroncong music mainly uses a system called chordal, namely Western musical concepts. This system uses a diatonic scale in the form of Solmization using Western harmony, namely chords I, IV, V, and II. Since the holding of the keroncong music group competition in Deca Park (City Park, now the Medan Merdeka area of Jakarta) Batavia, Concours (Het nieuws van den dag voor Nederlandsch-Indië, April 11, 1917) or, later called Bintang Radio, keroncong music began to be popular with a chordal system with arrangements using Western music presentation methods.

Keroncong, developed in the Surakarta area (Solo City), has its specialty. The specialty of Solo keroncong can also be seen in combining musical instruments. The use of instruments in Solo keroncong music is unique, and instruments such as the ukulele, guitar, flute, cello, and contrabass are used to create a distinctive sound. This Solo style of keroncong music shows genius in processing simple musical instruments through the

simplicity of the ukulele (cuk and cak) used to create a complex but light rhythm. Solo keroncong also has its uniqueness because it plays songs in Javanese. Javanese style is music typical of the Java region that uses Western instrumentation but presents it regionally, namely the slendro and pelog scales that imitate Javanese gamelan. Songs created by Maestros Gesang, Ismanto Sapari, Anjarany Any, WS Nardi, Dharmanto, and other Solo keroncong music artists have beautiful orchestration and instrumentation typical of the Javanese style.

Keroncong is also known for its rich harmony and melody. The character of Solo Keroncong's music is melancholic and beautiful. The harmony of keroncong music is famous for its softness and melodious melodies, which touch the listener's emotions. Its rhythm is also unique; with its distinctive rhythmic pattern, Solo style keroncong music can create a relaxed atmosphere full of emotion.

Keroncong music lyrics are often used to convey cultural messages and philosophical values. The poetic lyrics of keroncong music contain moral messages, teach the philosophy of life, or are expressions of love and longing. The beauty of its language reflects the genius of its poet and creator. For example, in the keroncong song "*Bahana Pancasila*" by Budiman BJ, the lyrics glorify Pancasila as the philosophy of the Indonesian nation.

Ultimately, keroncong, as a masterpiece of the Indonesian people, has become a national identity. Keroncong is a piece of music considered "truly Indonesian," reflecting the values of nationality, togetherness, and cultural identity that are the nation's pride. The aesthetics of keroncong music are eternal, and its beauty is conveyed by sound, structure, and emotions, making it a work of art that transcends time and is relevant in various eras. With these elements, keroncong is not just music but a masterpiece that illustrates the creative genius of humans in creating, adapting, and inheriting a culture of high value.

B. Keroncong Music as a Witness to Civilization

Keroncong music is considered a witness to civilization because it reflects the historical journey of the Indonesian nation. The long history of the growth of keroncong music began in the 17th century, when the Portuguese in Malacca came to Indonesia or the archipelago, especially in Kampung Tugu, Batavia. They have evolved and assimilated with the culture of the local community, which was initially a form of daily expression after they had been hunting, farming, raising animals, and fishing. They entertained themselves and each other by playing the music they could. At that time, their music was not yet called "keroncong" but was still music from their region as entertainment to relieve fatigue. They were enslaved Portuguese people captured by the Netherlands, a mixture of Portuguese with former colonies of India-Sri Lanka and Malacca, and local people who mostly came from Bengal and Malabar. The Dutch had to change their Portuguese-sounding Catholic names and replace them with new Dutch-style names that embraced Protestant Christianity, such as *Mardijkers* (a general term for freed former Asian and African slaves) people.

The development of keroncong music from the colonial era to the present shows that this genre can survive. Keroncong music has witnessed society's socio-political and cultural changes in every era. In the colonial era, keroncong witnessed the struggle of the Indonesian people against colonialism. Keroncong songs such as *Keroncong Kemayoran* were often used as symbols of national awakening and propaganda tools for the struggle. After Indonesia's independence, keroncong became part of national music and is now a symbol of national identity and a unifying tool.

Keroncong shows traces of extinct traditional music. Keroncong can be considered a "living archive" of traditional music that may have disappeared. Many elements of local traditional music are contained in keroncong, such as rhythmic and melodic patterns. Keroncong music often records certain events, values, or lifestyles in its lyrics, thus becoming valuable documentation of past ways of life. Keroncong also shows the tradition of a typical Indonesian ensemble based on harmony and cooperation. This shows the value of the collectivity of Indonesian society in the arts. Nowadays, keroncong has become an art form that connects the old and young generations. Keroncong proves how culture and tradition can survive amidst the current of modernity.

C. Keroncong Music has Traditional Values

In the lives of Indonesian people, keroncong music is still part of the socio-cultural tradition. Keroncong music is an integral part of the Indonesian cultural identity. The lyrics of the songs often contain values from the daily lives of the Indonesian people. Keroncong functions as a medium to preserve and convey local culture and traditions. The themes primarily found in keroncong songs are reflections of regional identity, love, food, love of the homeland, all kinds of cultural activities, advice, or local cultural behavior, both seriously and jokingly. For example, the legendary song "*Bengawan Solo*" was created by Gesang, a keroncong maestro from Solo; the lyrics describe the beauty of the Bengawan Solo River. Keroncong music still survives in Indonesia today because of its adaptability. Keroncong music in various regions is closely related to local culture. In Batavia, keroncong collided with the music culture of Gambang Kromong; in West Java, it blended with *jaipongan*; in Central Java, it was famous for its gamelan; in East Java, with *Ludruk*. Keroncong in Kalimantan and Sulawesi also blended with local music, culture, and traditions. While in Maluku, keroncong was influenced by the cheerful and passionate music of the Ambonese people. The cultural value of keroncong music reflects the musicality and style of its people.

Socially, keroncong music is often played at various events such as weddings, traditional ceremonies, and other social gatherings that help strengthen social relations and a sense of togetherness. At the beginning of the growth of keroncong music in Batavia, for the first time in 1880, musicians from Tugu village played their music outside their hometown. They played in the form of a stage parody with musical performances wearing European costumes and playing guitars with *Morisco* songs from Portugal by the native people. This music was often used for celebrations by Indo-Dutch crossbreed people. Lower-class Indo-European people favored keroncong music as a form of public entertainment. At that

time, popular entertainment was keroncong music because there were no competitors for other types of entertainment music. Keroncong music reached its golden age in the pre- and post-independence era.

Keroncong music is readily accepted because of its universal values. The songs performed in keroncong music are in Indonesian, using a standard diatonic scale. The types of songs from Western music played in keroncong rhythm/music also use Indonesian lyrics that various groups understand. This phenomenon strengthens the Indonesian nation because they have the same idiom, namely, keroncong music, and its language lyrics. Several times, envoys from Indonesia who participated in international music festivals performed keroncong music, which eventually became a characteristic of Indonesian music. This shows that keroncong music is a representation of Indonesia's national identity. This music symbolizes diversity and unity in difference, one of the basic principles of Indonesia's national motto, *Bhinneka Tunggal Ika* (unity in diversity).

ACKNOWLEDGMENT

The Ministry of Research, Technology, and Higher Education of Indonesia supported this work and funded it through the Domestic Collaborative Research Scheme in 2024.

REFERENCES

- [1] UNESCO, "Operational Guidelines for the Implementation of the World Heritage Convention," 2023. [Online]. Available: <https://whc.unesco.org/fr/orientations>
- [2] B. de-Miguel-Molina, V. Santamarina-Campos, M. de-Miguel-Molina, and R. Boix-Doménech, Eds., *Music as Intangible Cultural Heritage*. In *SpringerBriefs in Economics*. Cham: Springer International Publishing, 2021, doi: 10.1007/978-3-030-76882-9.
- [3] S. A. Putra, E. Ismariati, M. Hidayat, K. Setiagama, and Y. A. Nugroho, *Buku Penetapan Warisan Budaya Tak Benda Tahun 2020*, 1st ed., vol. 1. Jakarta: Direktorat Pelindungan Kebudayaan, Direktorat Jenderal Kebudayaan, Kementerian Pendidikan dan Kebudayaan, 2020.
- [4] D. Didit, "Interview/Wawancara Pengurus KAMKI di Kota Solo," Feb. 21, 2024, Solo.
- [5] J. S. Nye, "Public Diplomacy and Soft Power," *Ann Am Acad Pol Soc Sci*, vol. 616, no. 1, pp. 94–109, Mar. 2008, doi: 10.1177/0002716207311699.
- [6] J. S. Nye, "The Future of American Power: Dominance and Decline in Perspective," *Foreign Affairs*, vol. 89, no. 6, pp. 2–12, 2010, [Online]. Available: <http://www.jstor.org/stable/20788711>
- [7] D. I. Witarti and Y. C. Reza, "Pesan Perdamaian Pekan Olahraga Nasional (PON) XX Papua," *Jurnal Ilmu Komunikasi*, vol. 21, no. 1, pp. 113–131, May 2023, doi: 10.31315/jik.v21i1.7005.
- [8] D. I. Witarti and P. F. Ariyani, "Analisis Sentimen Berita Konser K-Pop Dalam Rangka Peringatan 50 Tahun Hubungan Bilateral Indonesia - Korea Selatan," Jakarta, Feb. 2024.
- [9] D. I. Witarti, A. Puspitasari, and A. Fithriana, "Revitalisasi Musik Keroncong dalam Pelaksanaan Diplomasi Budaya Indonesia ke Malaysia," Jakarta, Nov. 2023.
- [10] D. Sugiyanto and N. B. Aji, "Perkembangan Musik Keroncong di Richmod Virginia Amerika Serikat," *Keteg: Jurnal Pengetahuan, Pemikiran dan Kajian Tentang Bundy*, vol. 19, no. 2, pp. 141–154, May 2019, doi: 10.33153/keteg.v19i2.3080.
- [11] P. Kerr and G. Wiseman, *Diplomacy in a Globalizing World: Theories and Practices*, 1st ed. United Kingdom: Oxford University Press, 2017.
- [12] Q. Xu, Y. Xu, and C. Ma, "Analysis of contemporary value and influence of intangible cultural heritage based on online review mining," *PLoS One*, vol. 19, no. 12, p. e0315805, Dec. 2024, doi: 10.1371/journal.pone.0315805.
- [13] Q. Wang, "The Digitisation of Intangible Cultural Heritage Oriented to Inheritance and Dissemination under the Threshold of Neural Network Vision," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/6323811.
- [14] Y. Zhang and T. Ala, "Classification and Value Assessment of Sports Intangible Cultural Heritage Resources Combined with Digital Technology," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-0548.
- [15] Z. Zheng, "Tapping the Economic Potential and Optimizing Strategies of Intangible Cultural Heritage under Digital Transformation," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-3187.
- [16] G. Xu, Y. Meng, Z. Chen, X. Qiu, C. Wang, and H. Yao, "Research on Topic Detection and Tracking for Online News Texts," *IEEE Access*, vol. 7, pp. 58407–58418, 2019, doi: 10.1109/ACCESS.2019.2914097.
- [17] Y. J. Lee, S. E. Park, and S. Y. Lee, "Machine Learning-Driven Topic Modeling and Network Analysis to Uncover Shared Knowledge Networks for Sustainable Korea-Japan Intangible Cultural Heritage Cooperation," *Sustainability (Switzerland)*, vol. 16, no. 24, Dec. 2024, doi: 10.3390/su162410855.
- [18] B. Mudjiyanto, "Tipe Penelitian Eksploratif Komunikasi," *Jurnal Studi Komunikasi dan Media*, vol. 22, no. 1, pp. 65–74, Jun. 2018, doi: 10.31445/jskm.2018.220105.
- [19] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, vol. 5. United States: SAGE Publications, 2017.
- [20] R. A. Krueger and M. A. Casey, *Focus Groups: A Practical Guide for Applied Research*, 5th ed. United States: SAGE Publications, 2014.
- [21] H. J. Rubin and I. S. Rubin, *Qualitative Interviewing: The Art of Hearing Data*, 1st ed. United States: SAGE Publications, 2012.
- [22] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*. Springer, 2018.
- [23] S. Firman, W. Desena, and A. Wibowo, "Penerapan Algoritma Stemming Nazief & Adriani Pada Proses Klasterisasi Berita Berdasarkan Tematik Pada Laman (Web) Direktorat Jenderal HAM Menggunakan Rapidminer," *Syntax: Jurnal Informatika*, vol. 11, no. 2, pp. 10–21, Oct. 2022, doi: <https://doi.org/10.35706/syji.v11i02.7192>.
- [24] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [25] Q. Yang, "LDA-based Topic Mining Research on China's Government Data Governance Policy," *Social Security and Administration Management*, vol. 3, no. 2, pp. 33–42, 2022, doi: 10.23977/socsam.2022.030205.
- [26] S. K. Habibabadi and P. D. Haghghi, "Topic Modelling for Identification of Vaccine Reactions in Twitter," in *Proceedings of the Australasian Computer Science Week Multiconference*, in *ACSW '19*. New York, NY, USA: Association for Computing Machinery, Jan. 2019, pp. 1–10. doi: 10.1145/3290688.3290735.
- [27] H. Jelodar et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed Tools Appl*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.
- [28] J. L. Hung and K. Zhang, "Examining mobile learning trends 2003-2008: A categorical meta-trend analysis using text mining techniques," *J Comput High Educ*, vol. 24, no. 1, pp. 1–17, Apr. 2012, doi: 10.1007/s12528-011-9044-9.
- [29] S. Choi and J. Y. Seo, "An Exploratory Study of the Research on Caregiver Depression: Using Bibliometrics and LDA Topic Modeling," *Issues Ment Health Nurs*, vol. 41, no. 7, pp. 592–601, Apr. 2020, doi: 10.1080/01612840.2019.1705944.
- [30] N. A. Tresnasari, T. B. Adji, and A. E. Permanasari, "Social-Child-Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 14, no. 2, p. 179, Apr. 2020, doi: 10.22146/ijccs.54507.