Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

Density-Based Spatial Clustering, K-Means and Frequent Pattern Growth for Clustering and Association of Malay Cultural Text Data in Indonesia

Mustakim Mustakim*, Febi Nur Salisah

Department of Information Systems, Faculty of Science and Technology, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
Email: [1,*]mustakim@uin-suska.ac.id, [2]febinursalisah@uin-suska.ac.id
Correspondence Author Email: mustakim@uin-suska.ac.id
Submitted: **04/06/2025**; Accepted: **30/06/2025**; Published**: 30/06/2025**

Abstract−Several studies state the need to develop information technology to disseminate information related to culture in Indonesia. There are many similar studies but they still have weaknesses, one of which is that they do not use machine learning and intelligent computing. This research answers the challenges of previous researchers, namely developing machine learning-based learning applications using the Density-Based Spatial Clustering of Application Noise (DBSCAN) and Frequent Pattern Growth (FP-Growth) algorithms. The results of the modeling of the two algorithms are deemed to still require improvement in the future, as it is proven that DBSCAN does not yet have optimal validity. So in this research, one of the comparison algorithms is used, namely K-Means Clustering, with a better evaluation than DBSCAN. The modeling results were implemented into mobile programming as a cultural learning application in Indonesia, especially Riau Malay Culture, the black box testing results had an accuracy of 100% and the User Acceptance Test (UAT) was 86%. Thus, it is concluded that this application can be used effectively and efficiently for general users.

Keywords: DBSCAN; FP-Growth; K-Means; Riau Malay

1. INTRODUCTION

Indonesia is the largest archipelagic country in the world, there are various ethnicities and cultures in it. There are various ethnic and cultural characteristics in Indonesia, each region has differences. The cultural characteristics of each region include customs, regional clothing, traditional houses, regional arts, regional languages, and regional specialties [1]. In the current era of globalization, cultural values in society seem to be increasingly extinct and traditional art forms are dying in the archipelago. This condition results in cultural heritage and local wisdom in Indonesia becoming increasingly extinct [2]. One of the cultures that characterizes Islam in Indonesia is Malay culture.

Riau Province, known as Bumi Lancang Kuning or Bumi Melayu, is the largest historical site of the Malay kingdom in Indonesia and is an area that has a variety of cultures that are known throughout the world [3]. Riau Malay culture is the result of the thoughts, feelings, and work of Malay people who speak Malay, have Malay customs and are Muslim [4]. Law of the Republic of Indonesia Number 5 of 2017 states that Malay culture in Riau Province makes a huge contribution to national culture [5], likewise the religion and culture of the archipelago is currently influenced by Malay culture. However, in reality, the relationship between state power and customs is currently not balanced [6], one of which is the continued recognition of Malay culture by other countries, as a result of the current lack of generational learning.

Several studies have stated that it is necessary to develop information technology to disseminate information related to culture in Indonesia. Such as research conducted by Muhammad Juanda Saputra for Acehnese culture [7], Abdurahman Dayat for Papuan culture [8], Yulisman for Riau culture [3], and Ahmad Suryadi for Introduction to tribes in Indonesia [9]. The weakness of these studies lies in two important things in applying technological aspects, namely Intelligent Computing Technology (*Computational Intelligence*) and Machine Learning, which were not implemented in the research so they tend to be limited to ordinary applications that do not recommend anything.

Machine Learning technology is currently developing rapidly and has been implemented on many platforms, one of which is mobile-based [10]. Machine Learning is a field of science that applies several data and mathematical concepts in modeling, discovering patterns, and determining new concepts in solving problems [11]. One part of Machine Learning is text-based learning technology, or often known as Text Mining [12]. Modeling is carried out in learning by calculating several text documents based on distances and rules known as TF-IDF [13], processing by applying a certain algorithm to provide better validity or accuracy [14], and producing conclusions according to the calculation of a certain method [15].

The implementation in the case of text mining for Malay Culture is being able to measure how accurate it is based on association rules in grouping data. Each text document is grouped first using the Density-Based Spatial Clustering of Application Noise (DBSCAN) algorithm and then formulated in Association Rules using the Frequent Pattern Growth (FP-Growth) algorithm. These two algorithms were carried out simultaneously in one case in Mustakim and Ulya Khairunnisa's research for modeling sales data in supermarkets in Pekanbaru City. The result was that combining these two methods had maximum performance in the data matching aspect [16]. Apart from that, DBSCAN is a clustering-based algorithm that has good performance in the case of large data [17], independently by forming groups based on its outlayer [18], and can handle data that has large missing values [19]. This algorithm with

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

high performance and complexity can be used to develop a Digital Learning concept for learning both in schools, universities, and learning in general.

Machine Learning algorithms in mobile applications are applied in this research by conducting better experiments and combining two different algorithms into one case. The accuracy of the modeling carried out can recommend text documents in applications that have the closest level of similarity and have a higher relationship than the hundreds of documents accessed. So that users can easily identify directly both from the article and from the Malay culture sub-section. This application is expected to be a breakthrough for community learning related to Malay culture in Indonesia. The novelty of this research includes experimental aspects and a combination of Machine Learning algorithms with the best formulation applied to DLS, as an intelligent machine in mapping Malay cultural learning in Indonesia, which until now has never been done by other researchers. This research aims to model the Density-Based Spatial Clustering of Application Noise (DBSCAN) algorithm and Frequent Pattern Growth (FP-Growth) Association Rules as an intelligent machine in mapping text documents in the case of Malay Culture in Indonesia. Next, build a Digital Learning System (DLS) based on the best modeling algorithm DBSCAN and FP-Growth to be implemented in a mobile application platform. This research contributes significantly by offering a novel, automated approach to clustering and associating cultural texts, enabling smarter and more accessible learning of Malay culture through digital means. With this application, the wider community can be helped to learn and understand Malay culture correctly, based on structured, relevant, and contextually accurate information.

2. RESEARCH METHODOLOGY

2.1 Materials and Methods

This research process is divided into four main stages, namely (1) User Requirements and Data Collection; (2) data pre-processing, experimentation and data modeling using two Machine Learning algorithms, namely DBSCAN for grouping or mapping data and FP-Growth for determining association relationships in data; (3) development of mobile applications using the Android and web platforms, built by applying the concept of Object Oriented Programming (OOP) with a waterfall model approach; (4) application testing and user training. In general, the research methodology can be seen in Figure 1, while the relevant research map related to this study can be seen in Figure 2.

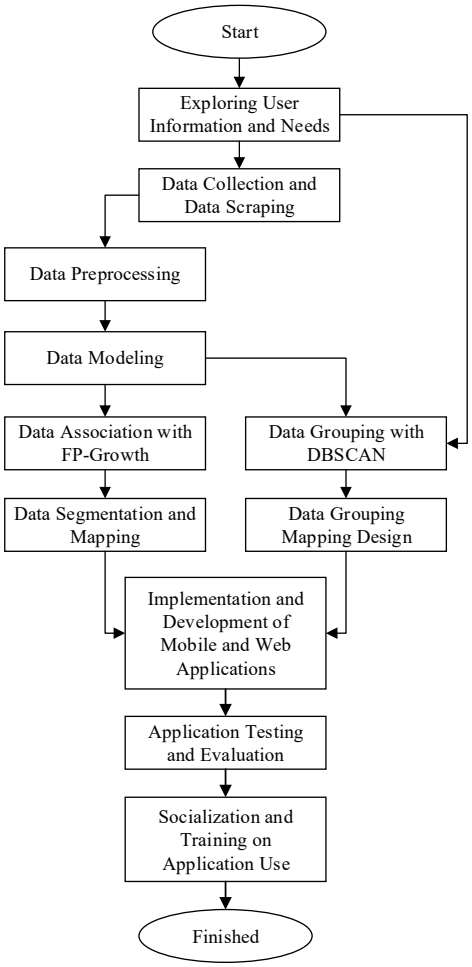


Figure 1. Research Methodology

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

The main attributes and categories used in this research are (1) Oral Traditions, (2) Manuscripts, (3) Customs, (4) Rites, (5) Traditional Knowledge, (6) Traditional Technology, (7) Art, (8) Language, (9) Folk Games and (10) Traditional Sports. This attribute is used as a guide and reference for collecting text data and building structures in mobile applications. In general, the research methodology can be seen in Figure 1.

Machine Learning algorithm modeling taken directly from the internet in all sources related to Malay culture in Indonesia. This data will later be used for the text mining analysis process as mapping/grouping and searching for association relationships between data. This data retrieval process uses scrapping techniques using the Python programming language. The corpus and documents collected in a large database are then cleaned to be integrated with several tabular data originating from other data sources.

Seeing the large number of previous studies that have been carried out in the case of developing technology to study culture in Indonesia, such as (1) introduction to the culture of WITA provinces [20]; (2) introduction to Belu district culture [21]; (3) introduction of typical Toraja culture [22]; (4) designing Android-based game applications for Riau Malay traditions [23]; (5) introduction to the culture of Riau province [3]; (6) introduction of Tangerang cultural heritage [24]; (7) Introduction to Ethnicity and Culture [9]; (8) Android-based history of Acehnese culture [7]; (9) Design of Augmented Reality Based Application for Introduction to Typical Papuan Culture [8]. Each of these studies has advantages, but there are also many weaknesses in this research, one of which is related to the non-implementation of Intelligent Systems, Machine Learning, and Text-Based Analysis. Apart from that, the platforms that have been built have not yet adopted the Digital Learning System (DLS) concept, which until now has been integrated with the Digital Literacy concept.

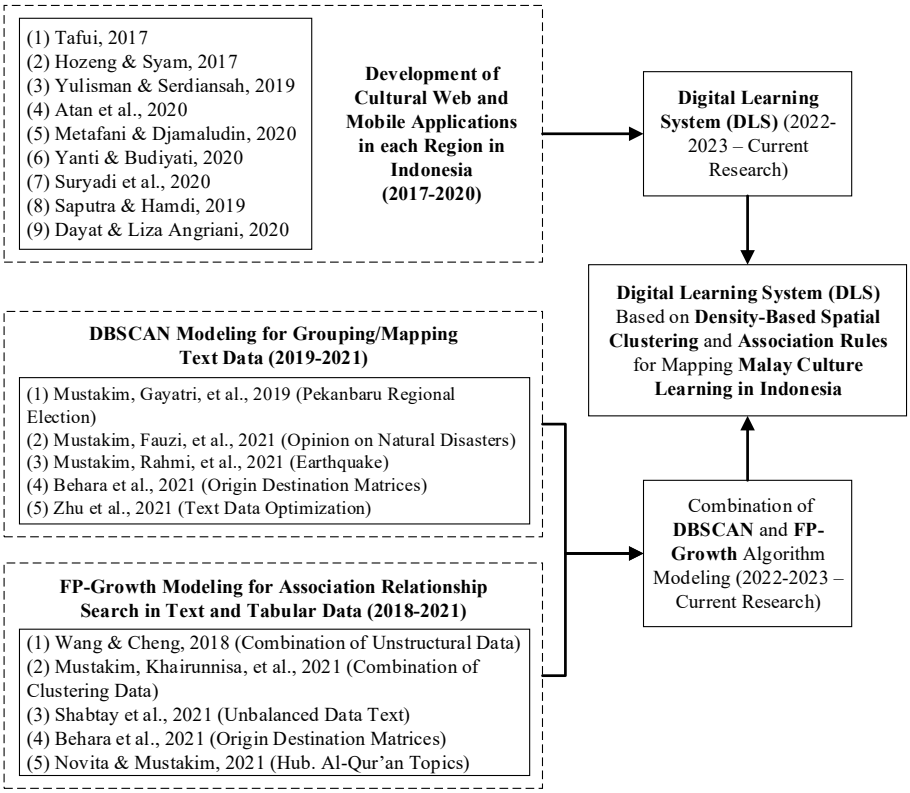


Figure 2. Relevant Research

DBSCAN is an algorithm part of Unsupervised Learning often called Clustering. Research related to DBSCAN in the case of text modeling or text mining for analyzing events sourced from text documents and social media was carried out by Mustakim and friends in 2019-2021 in the case of mapping candidates for Mayor of Pekanbaru [25], comparing K-Means and DBSCAN for public opinion about natural disasters in Indonesia, DBSCAN has high cluster validity [15], and earthquake modeling in Indonesia is collaborated with the Principal Component Analysis (PCA) algorithm [18], and this algorithm also has good performance to map origin-destination matrices compared with the Fuzzy C-Means and K-Medoid algorithms [26]. In terms of application to searches in web applications and other platforms, DBSCAN has very optimal performance with very low complexity values [27]. It can be concluded that for the case of text documents, the DBSCAN algorithm has good performance, high cluster validity, and low algorithm complexity. So, in its application for Malay culture learning mapping applications in Indonesia, DBSCAN is very suitable to be one of the intelligent machines to be applied in this application.

FP-Growth is an algorithm that is widely applied to text document cases compared to Apriori and ECLAT. Research related to FP-Growth was carried out by Rice Novita and Mustakim to accurately map text documents in the Al-Qur'an related to the topics contained therein [28]. In other cases, FP-Growth is also able to provide

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

recommendations for the best transactions based on the relationship between items in a retail company's customer data combined with a clustering algorithm [16]. Lior Shabtay and friends also stated that FP-Growth was able to overcome some unbalanced data, especially with regard to very large data [29]. The reliability of this algorithm is also able to monitor data in a system that is correlated with each other, to form stable and constant data [30]. These studies prove that the reliability of FP-Growth Association Rules will certainly support the process of forming clusters and mapping to add the best performance to DBSCAN in terms of data combination.

3. RESULT AND DISCUSSION

The results and analysis in this report are divided into 3 main parts, namely: (1) pre-processing, Modeling, and Experimentation using Machine Learning algorithms, (3) User Requirements and Mobile application development, and (3) Application testing and training for users.

3.1 Pre-Processing, Modeling, and Experiments Using Machine Learning Algorithms

In data mining modeling, several datasets are needed for the algorithm experiment process. This research uses 2 algorithms, namely DBSCAN as a grouping algorithm and the FP-Growth algorithm as a word-matching algorithm.

The first process carried out was collecting a dataset, in this case, the dataset used was unstructured data from the internet in the form of data relating to Malay culture, the amount of data used consisted of 14 categories and 156 data. As an example, several datasets from the web scapping process are given which are shown in Table 1.

Table 1. Sample Datasets

No	Category	Main Name of Culture	Short Description of Corpus Data
1	Legenda Cerita Rakyat	Legenda Siak Sri Indrapura	Siak Sri Indrapura adalah sebuah kerajaan di Riau yang memiliki banyak cerita rakyat. Salah satu legenda terkenal menceritakan tentang pendirian kerajaan ini oleh Raja Kecik, seorang pemuda yang bijaksana dan berani.
2	Legenda Cerita Rakyat	Teruskan Kebawah	Sang Nila Utama adalah seorang pahlawan legendaris yang konon merupakan pendiri Singapura. Menurut legenda, ia datang ke pulau tersebut dan melihat seekor singa, yang kemudian menjadi nama Singapura. Kisah ini memiliki keterkaitan dengan wilayah Riau karena terdapat hubungan historis antara Riau dan Singapura.
3	Tradisi Lisan	Pentas Wayang Kulit	Legenda ini mengisahkan seorang putri bernama Mayang Mangurai yang memiliki keterampilan dalam menenun. Cerita ini mengangkat nilai-nilai seperti kebijaksanaan dan keterampilan kerajinan.
…	…	…	…
156	Tokoh Melayu Riau	Raja Ali Haji	Raja Ali Haji adalah seorang penyair dan ulama terkenal dari Riau yang terkenal karena karyanya, "Gurindam Dua Belas." Kisah hidupnya dan kontribusinya dalam bidang sastra dan agama menjadi subjek legenda.

The main corpus used in research is based on 14 categories which will later be grouped using the DBSCAN algorithm. Table 2 shows the total corpus of main cultural names used as a sample dataset.

Table 2. Number of Corpus Sample Datasets

No	Category	Amount of Data
1	Legenda Cerita Rakyat	8
2	Tradisi Lisan	13
3	Manuskrip	7
4	Adat Istiadat	15
5	Ritus budaya	13
6	Pengetahuan Tradisional	13
7	Teknologi Tradisional	14
8	Seni Kebudayaan	8
9	Bahasa melayu	7
10	Permainan Rakyat	17
11	Olahraga Tradisional	6
12	Kerajaan Melayu	6
13	Raja Kerajaan Melayu Riau	8

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

No	Category	Amount of Data
14	Daerah Bekas Kerajaan	9
15	Tokoh Melayu Riau	12
	Amount	156

From Table 1 and Table 2, the data underwent pre-processing through several stages, such as cleaning, tokenising, filtering, and stemming. This process was carried out to convert the original data into several parts of data that would be used for the data clustering process. Usually, the data would be converted into numerical values from the TF-IDF process. Data pre-processing as a sample in record 1 is shown in Table 3.

Table 3. Pre-Processing Data

Step	Result
Cleaning	siak sri indrapura adalah sebuah kerajaan di riau yang memiliki banyak dongeng rakyat salah satu legenda terkenal menceritakan tentang pendirian kerajaan ini oleh raja kecik seorang pemuda yang bijaksana dan pemberani (hapus tanda baca, huruf kecil semua)
Tokenizing	[siak, sri, indrapura, adalah, sebuah, kerajaan, di, riau, yang, memiliki, banyak, dongeng, rakyat, salah, satu, legenda, terkenal, menceritakan, tentang, pendirian, kerajaan, ini, oleh, raja, kecik, seorang, pemuda, yang, bijaksana, dan, pemberani]
Filtering	[siak, sri, indrapura, kerajaan, riau, memiliki, dongeng, rakyat, legenda, terkenal, menceritakan, pendirian, kerajaan, raja, kecik, pemuda, bijaksana, pemberani] (stopword seperti: adalah, sebuah, di, yang, salah, satu, tentang, ini, oleh, seorang, dan → dihapus)
Stemming	[siak, sri, indrapura, raja, riau, milik, dongeng, rakyat, legenda, kenal, cerita, diri, raja, kecik, muda, bijak, berani] (menggunakan stemmer bahasa Indonesia, misal dari Sastrawi)

After passing through the preprocessing stage, the data is then weighted using the TF-IDF method. The weight value of a term increases with its frequency of occurrence. TF-IDF calculations are performed using the Python programming language with the Scikit-Learn library. The results of the TF-IDF process can be seen in Table 4.

Table 4. TF-IDF

Doc/ Term	bijak	brave	cerita	dongeng	indrapura	kecik	…	raja
Doc1	0.00	0.67	0.00	0.45	0.00	0.00	…	0.00
Doc2	0.00	0.00	0.00	0.00	0.00	0.25	…	0.00
Doc3	0.00	0.75	0.00	0.00	0.00	0.33	…	0.00
Doc4	0.00	0.00	0.00	0.00	0.20	0.40	…	0.00
Doc5	0.00	0.82	0.46	0.00	0.00	0.00	…	0.00
Doc6	0.00	0.00	0.00	0.00	0.77	0.00	…	0.00
Doc7	0.00	0.00	0.22	0.00	0.00	0.52	…	0.00
Doc8	0.00	0.39	0.00	0.00	0.00	0.00	…	0.00
…	…	…	…	…	…	…	…	…
Doc156	0.00	0.00	0.00	0.00	0.61	0.00	…	0.00

The labelling process is carried out automatically using the Indonesian sentiment dictionary (InSet). In practice, each word is compared with entries in the dictionary. If a match is found, the polarity value of that word is taken and added to the total.

3.2 Experiment with the DBSCAN Algorithm

Experiments with the DBSCAN algorithm, this process is carried out to model the best data from a series of experiments carried out. The experiment produces cluster values, cluster validity, and non-space clusters. Where the results of the best cluster will later be used as mapping in the application development process, of course, the validity value of the best cluster is calculated first. Clusters will be formed automatically by the DBSCAN algorithm machine with a certain formulation and processed using the Python programming language with Google Colab. The results of the experiment using DBSCAN can be seen in Table 3 and Table 4. Evaluation using the Davies-Bouldin Index (DBI) and Silhouette Score/ Index (SI). The parameters used for the DBSCAN algorithm are eps = 0.3 and min_samples = 5. Eps is the maximum radius of the nearest neighbour for two points to be considered in the same cluster, while min_samples is the minimum number of points (including the centre point) required to form a core point. This yields the values shown in Table 5 and Table 6.

Table 5. Evaluation Results of Experiment 1 (7 Clusters) on the DBSCAN Algorithm

Evaluation	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Outlayer
Davies-Bouldin Index (DBI)	6,594	6,710	6,171	6,381	7,376	7,599	10,350
Silhouette Score/ Index (SI)	0.002	0.002	0.002	0.006	0.004	0.008	0.006

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

Table 6. Evaluation Results of Experiment 2 (5 Clusters) on the DBSCAN Algorithm

Evaluation	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Outlayer
Davies-Bouldin Index (DBI)	8,611	5,761	5,923	5,740	13,770
Silhouette Score/ Index (SI)	0.002	0.002	0.002	0.006	0.006

Because this research has low validity results when using DBSCAN, cluster experiments were also carried out by trying out the K-Means Clustering algorithm. The experimental results used 3 validity clusters, namely the Davies-Bouldin Index (DBI), Silhouette Score/Index (SI), and Elbow. The results of the elbow method can be shown in Figure 3.

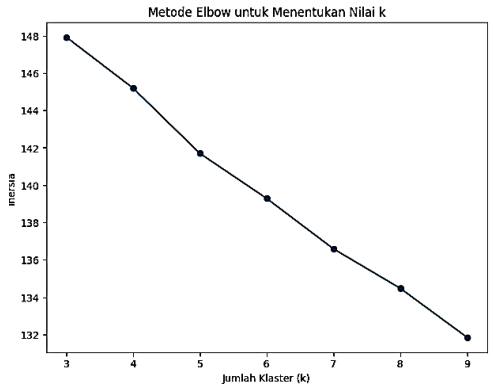
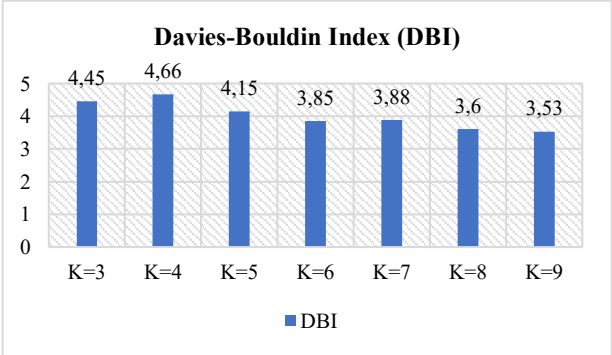


Figure 3. Evaluation Using the Elbow Method

Based on the elbow method calculation, cluster calculations were then carried out using K-Means. The k-value trials used were 3, 4, 5, 6, 7, 8, and 9. Each trial was evaluated using the Davies-Bouldin Index (DBI) and Silhouette Index (SI). The evaluation results of DBI and SI are shown in Figures 4 and Figure 5 respectively.



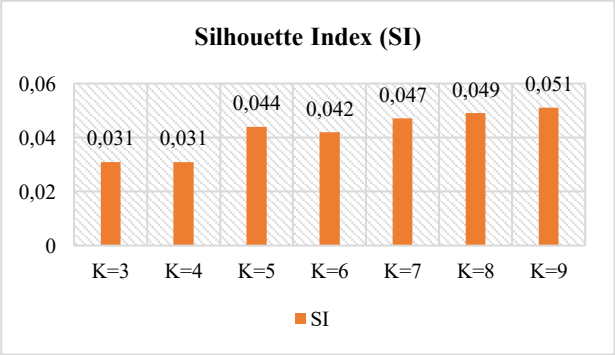


Figure 4. Evaluation with DBI

Figure 5. Evaluation with SI

Based on the DBI evaluation, the most optimal k value that is close to zero is k=9. Meanwhile, in the SI evaluation, the most optimal k value is k=9. So it is concluded that clustering in this algorithm uses k=9. Figure 6 is the percentage of cluster members obtained from the best evaluation scores. The results of this process are dominated by cluster 3, can be view as Figure 6. Next, data visualization was carried out for each cluster using the Wordcloud shown in Figure 7 as a sample.

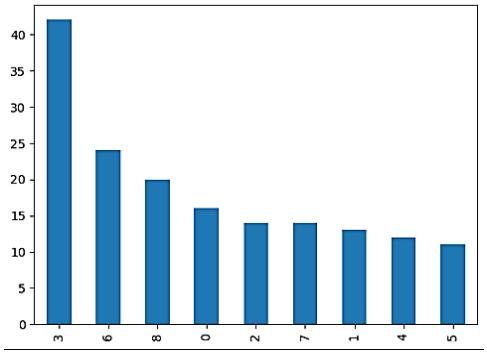


Figure 6. Percentage of Cluster Member Dominance

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

Figure 7. Worldcloud in Cluster 0

3.3 Experiment with FP-Growth

In experiments with FP-Growth, this process is carried out to model association relationships between data from clusters that have been formed. This FP-Growth engine will maximally model the relationships between attributes based on words in a sentence in text data. The results of this modeling will form support and confidence values with predetermined parameters. Its implementation in the application becomes a measure of sensitivity to determine the relationship of each word or sentence entered into the application. The results of the three experiments can be shown in Tables 7, 8, 9 and 10.

Table 7. Results of Support Values for Words that Often Appear in Experiment 1

No	Frequently Appearing Words	Amount	Support
1	Melayu Riau	521	21.7%
2	Wilayah	422	18.2%
3	Raja	368	9.3%
4	Riau	355	20.5%
5	Salah	276	3.3%
6	Lingga	182	17.4%
7	Pulau Sengat	180	8.1%
8	Provinsi Riau	177	17.7%
9	Raja Haji	124	12.4%
10	Bekas Wilayah	119	7.4%

Table 8. Results of Support Values for Words that Often Appear in Experiment 2

No	Frequently Appearing Words	Amount	Support
1	Tradisional	447	11.7%
2	Melayu Riau	319	23.6%
3	Musik	307	9.8%
4	Tari Zapin	304	19.6%
5	Upacara Adat	267	9.1%
6	Rebana	260	4.4%
7	Gendang	241	4.6%
8	Budaya Melayu	238	20.5%
9	Musik	189	2.8%
10	Rebana	174	3.7%

Table 9. Results of Support Values for Words that Often Appear in Experiment 3

No	Frequently Appearing Words	Amount	Support
1	Siak Sri Indrapura	449	17.3%
2	Siak	312	19.8%
3	Raja Siak	310	9.1%
4	Raja Melayu	287	23.0%

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

No	Frequently Appearing Words	Amount	Support
5	Istana	276	3.3%
6	Peran	189	16.2%
7	Masjid	171	7.3%
8	Abad 19	153	12.1%
9	Kota Siak	114	9.3%
10	Masjid Sejarah	103	6.4%

So from these three tables, we can take some of the words that appear most frequently and have combinations which can be shown in Table 10.

Table 10. Results of the Largest Support Value in the Three Experiments

No	Frequently Appearing Words	Amount	Support
1	Melayu Riau	840	45.3%
2	Raja Melayu	287	23.0%
3	Riau	355	20.5%
4	Budaya Melayu	238	20.5%
5	Siak	312	19.8%
6	Tari Zapin	304	19.6%
7	Wilayah	422	18.2%
8	Provinsi Riau	177	17.7%
9	Lingga	182	17.4%
10	Siak Sri Indrapura	449	17.3%

With several final processes, the rules for concluding related words consist of 65 rules. As a sample in this study, the 10 highest rules from the FP-Growth algorithm process are displayed, which are shown in Table 11. Furthermore, the combination of 3 items or 3 words that often coincide can be seen in Table 12.

Table 11. Combination of 2 Words in FP-Growth

No	Combination of 2 Words	Support	Confidence
1	IF Melayu Riau THEN Riau	13.3%	57.9%
2	IF Melayu Riau THEN Budaya Melayu	10.1%	52.3%
3	IF Melayu Riau THEN Siak Sri Indrapura	18.4%	48.3%
4	IF Riau THEN Tari Zapin	17.9%	35.2%
5	IF Riau THEN Budaya Melayu	10.4%	44.2%
6	IF Budaya Melayu THEN Siak Sri Indrapura	11.3%	58.1%
7	IF Budaya Melayu THEN Profinsi Riau	9.8%	38.3%
8	IF Provinsi Riau THEN Tari Zapin	9.0%	48.9%
9	IF Siak Sri Indrapura THEN Raja Melayu	13.8%	59.8%
10	IF Siak Sri Indrapura THEN Melayu Riau	17.4%	37.2%

Table 3. Combination of 3 Words in FP-Growth

No	Combination of 2 Words	Support	Confidence
1	IF Melayu Riau THEN Riau AND Budaya Melayu	17.5%	44.6%
2	IF Riau THEN Budaya Melayu AND Provinsi Riau	9.8%	32.9%
3	IF Budaya Melayu THEN Siak Sri Indrapura AND Melayu Riau	14.6%	31.4%
4	IF Siak Sri Indrapura THEN Melayu Riau AND Raja Melayu	16.6%	29.8%

Intepretation of combination of 3 words is Based on the results of association modelling, several rules were found that indicate a strong correlation between keywords in the context of Riau Malay culture. The first rule indicates that when the term ‘Malay Riau’ appears, there is a 44.6% probability that the terms “Riau” and ‘Malay Culture’ will also appear together, with a total occurrence rate (support) of 17.5%. This indicates that ‘Malay Riau’ is closely associated with the cultural identity and geographical region of Riau. Furthermore, the word ‘Riau’ itself is also often associated with ‘Malay Culture’ and ‘Riau Province,’ although with a lower support rate of 9.8% and a confidence level of 32.9%. This shows that discussions about Riau are generally in the context of culture and regional administration.

The third rule shows that when the term ‘Malay culture’ is mentioned, there is a 31.4% chance that ‘Siak Sri Indrapura’ and ‘Riau Malay’ will also appear, with a support rate of 14.6%. This confirms that Malay culture is often associated with historical locations such as Siak Sri Indrapura and the Riau Malay ethnic group. Finally, the fourth rule shows that the appearance of the term ‘Siak Sri Indrapura’ is followed by the terms ‘Riau Malay’ and ‘Malay

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

King' 29.8% of the time, with a support rate of 16.6%, indicating that this historic city is closely linked to the Malay kingdom and the Malay cultural heritage in Riau. These four rules collectively provide a deeper understanding of how Malay cultural keywords are interconnected and can be leveraged for cultural-based learning systems or digital recommendations. If simplified, it would become:

a. When people talk about Melayu Riau, they often also refer to Riau and Budaya Melayu, with a 44.6% chance based on the data.
b. The topic Riau is frequently associated with Budaya Melayu and Provinsi Riau, although less dominant, with a 32.9% confidence level.
c. If the discussion involves Budaya Melayu, there is a 31.4% likelihood that Siak Sri Indrapura and Melayu Riau will also be mentioned.
d. When Siak Sri Indrapura appears in the context, it is often followed by references to Melayu Riau and Raja Melayu, with a 29.8% confidence level.

3.4 Mobile Application Development with the Android Platform

This section describes the overall design and development process of the mobile application. The development focuses on two critical aspects: how the application is designed with User Experience (UX) principles in mind, and how it is implemented using a suitable programming language. A user-centric approach is adopted to ensure the application's interface is intuitive, visually appealing, and aligned with users' expectations and behavior patterns. The mobile application is built using a cross-platform framework that allows deployment on both Android and iOS devices, ensuring broader accessibility and user reach. The development is carried out using object-oriented programming (OOP) principles, promoting modularity, reusability, and ease of maintenance. Each component of the application, such as user interfaces, backend logic, and data management layers, is encapsulated in individual classes and objects, making the system easier to understand, test, and scale in future iterations.

In terms of development methodology, the project adopts the Waterfall model, which consists of a linear sequence of phases: requirement analysis, system design, implementation, testing, deployment, and maintenance. This model is chosen due to its structured nature, which is suitable for clearly defined requirements and a stable project scope. The systematic approach helps in thorough documentation, better planning, and clear progress tracking throughout the development cycle.

The database used in this application is Firebase Realtime Database, a cloud-hosted NoSQL database that stores data in JSON format. It enables real-time data syncing between users and devices, which is essential for the application's dynamic and interactive features. Firebase also supports user authentication, secure data access, and easy integration with other Google services, making it a reliable choice for scalable mobile app development.

After development, the application undergoes a BlackBox Testing and User Acceptance Testing (UAT) phase to evaluate whether it meets the initial requirements and is ready for deployment. During this stage, selected end users interact with the application and provide feedback based on their experience. The test focuses on usability, functionality, responsiveness, and overall satisfaction. The results of the UAT show that 86% of the test participants are able to use the application effectively, indicating that the app is user-friendly, accessible, and ready for public release. This positive outcome reflects the effectiveness of the UX-driven design process and confirms that the features implemented meet users' expectations.

In conclusion, the development of this mobile application combines thoughtful design based on UX principles, structured programming using OOP, and a disciplined development approach via the Waterfall model. The use of Firebase as the backend solution ensures real-time, secure, and scalable data handling, while the high UAT result demonstrates the application's readiness and potential to deliver a meaningful and valuable user experience.

3.5 Interface Design and Testing

Interface design is the design of the main menu display on a system or application that has functions that can be carried out by the user. The interface design must include a good display that is easy to understand and display menus that are easy to understand. Some examples shown in designing the interface are the Homepage which is the main page of an application, a list of Riau Malay cultural information, detailed cultural information, a list of folklore, details of folklore and locations of Malay cultural history, as well as several other sections.

This application was built using the JavaScript programming language with MySQL as the database. This Malay cultural application uses the React Native framework to create an Android-based application. In the implementation stage, there are 2 main parts discussed, namely database implementation (given several database samples) and user implementation often referred to as coding. Next, the second part is the implementation of coding. In the analysis carried out, this process has passed user requirements to parties directly related to the research object. Figure 8 is an implementation image of the application coding.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

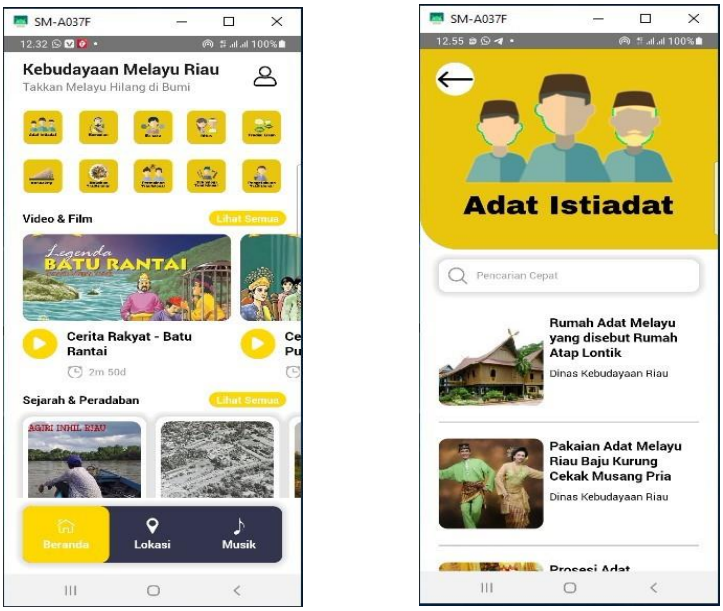


Figure 1. Application Implementation/Coding

On the application, black box testing has been carried out for users, it can be stated that the application is successful and running successfully. In the UAT testing that has been carried out on users, it can be stated that the percentage in application testing on users obtained a value of 86%, this can state that the Android-based application for introducing Riau Malay culture can be used well by users.

The final stage of application implementation is testing. Testing is conducted using two types of tests: Blackbox and User Acceptance Test (UAT). UAT testing is a type of system testing directly conducted by users, aiming to determine whether the developed system can serve as a solution for the system's objectives and simplify previous work processes. UAT is a test conducted by users to produce a document serving as evidence of whether the developed system is acceptable to users. If the test results are deemed to meet user needs, the application can be implemented. UAT testing was conducted by asking several questions to employees of the Riau Province Cultural Office, Riau cultural figures, and students acting as users. This testing involved 5 employees of the Riau Province Cultural Office, 5 Riau cultural figures, 5 Malay cultural artists, and 5 students as users. The UAT results were evaluated using 5 categories: SS (Very Suitable), S (Suitable), KS (Less Suitable), TS (Not Suitable), and STS (Very Not Suitable). The following is a detailed breakdown of the results.

The questionnaire distributed to users can be scored for each assessment aspect using the indicators listed in the Appendix. The UAT calculation results can be seen in Table 13.

Table 13. User Aceptance Test

Respondent	Question Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Respondent 1	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 2	5	5	5	5	4	5	5	4	4	4	4	4	4	4	5
Respondent 3	5	5	5	5	4	5	5	5	4	4	4	5	4	4	5
Respondent 4	5	5	5	5	4	5	5	5	5	5	5	5	4	4	5
Respondent 5	5	5	5	5	4	4	5	5	5	4	4	5	5	4	5
Respondent 6	5	4	4	5	4	5	5	5	5	5	5	5	5	3	5
Respondent 7	5	5	4	5	4	5	5	5	4	4	4	5	4	4	5
Respondent 8	5	5	5	5	4	5	5	5	5	5	5	5	4	4	5
Respondent 9	5	5	5	5	4	5	5	4	5	4	5	4	4	4	5
Respondent 10	5	5	5	5	4	5	5	5	5	5	5	5	4	4	5
Respondent 11	5	5	5	5	4	4	4	5	5	4	4	5	5	4	5
Respondent 12	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 13	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 14	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 15	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Respondent 16	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 17	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Respondent 18	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Respondent 19	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

Respondent	Question Number														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Respondent 20	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4
Total	90	89	88	90	73	88	89	88	87	84	85	88	83	78	90
Percentage	90	89	88	90	73	88	89	88	87	84	85	88	83	78	90

Blackbox testing has been conducted on the application to ensure that it runs successfully and smoothly for users. Based on the UAT testing conducted on users, the application achieved a score of 86% in user testing, indicating that the Android-based Riau Malay culture recognition application can be used effectively by users.

4. CONCLUSION

The research shows that clustering using the DBSCAN algorithm was tested with two scenarios: 7 clusters and 5 clusters. However, the evaluation results using DBI and SI scores were not optimal. Therefore, the K-Means algorithm was used as an alternative and produced better clustering results. K-Means was then chosen for grouping words before the association process. The association analysis used the FP-Growth algorithm and produced good rules with high support and confidence values. The most frequently searched words related to Malay culture were Melayu Riau, Raja Melayu, Riau, and Budaya Melayu, each with support values above 20%. From the top 10 words, the best association rules found were: IF Melayu Riau THEN Riau and IF Melayu Riau THEN Budaya Melayu, along with a combined rule: IF Melayu Riau THEN Riau AND Budaya Melayu. The rule "IF Melayu Riau THEN Riau" means that when users search for Melayu Riau, they are also likely interested in Riau. For users, this helps guide them to related content. For content managers, it can be used to automatically recommend relevant materials. These kinds of rules support the idea of a smart system that maps learning content related to Malay culture based on user interest and keyword patterns. The Digital Learning System (DLS) developed in this study also performed well in testing. The User Acceptance Test (UAT) scored 86%, the Blackbox test showed 100% functionality, and simulations on different devices ran smoothly. Overall, the system met expectations. However, the DBSCAN clustering could still be optimized or replaced with classification methods for future development, especially to support learning and preserving Riau Malay culture.

REFERENCES

[1] M. P. Sari and A. R. Hidayatulloh, "Pengenalan Kebudayaan Indonesia melalui Fotografi pada Akun Instagram 'KWODOKIJO,'" *Edsence J. Pendidik. Multimed.*, vol. 2, no. 2, pp. 111–120, 2020, doi: 10.17509/edsence.v2i2.27460.
[2] N. D. B. Setyaningrum, "Budaya Lokal Di Era Global," *Ekspresi Seni*, vol. 20, no. 2, p. 102, 2018, doi: 10.26887/ekse.v20i2.392.
[3] Y. Yulisman and S. Serdiansah, "Aplikasi Pengenalan Kebudayaan Provinsi Riau Berbasis Android," *J. Teknol. Sist. Inf. dan Apl.*, vol. 2, no. 3, p. 79, 2019, doi: 10.32493/jtsi.v2i3.3294.
[4] G. Riau, "Peraturan Daerah Provinsi Riau Nomor 9 Tahun 2015 Tentang Pelestarian Kebudayaan Melayu Riau," pp. 1–28, 2015.
[5] 2017 UU Nomor 5 Tahun, "UU 5 tahun 2017 tentang Pemajuan Kebudayaan," *2017, UU Nomor 5 Tahun*, p. 57, 2017.
[6] H. Wazni; Zulfa, "Relasi Kuasa Negara dan Adat dalam Mengembangkan Pariwisata Budaya Melayu Kabupaten Siak," *J. PolGov*, vol. 3, no. 2, pp. 361–392, 2021.
[7] M. J. Saputra and N. Hamdi, "Rancang Bangun Aplikasi Sejarah Kebudayaan Aceh Berbasis Android Studi Kasus Dinas Kebudayaan Dan Pariwisata Aceh," vol. 5, no. 2, pp. 147–158, 2019.
[8] A. R. Dayat and Liza Angriani, "Perancangan Aplikasi Pengenalan Kebudayaan Khas Papua Berbasis Augmented Reality," *JISKa*, vol. 5, no. 1, pp. 42–55, 2020.
[9] A. Suryadi, N. M. Rosa, and E. Subandriyo, "Perancangan Aplikasi Pengenalan Suku Dan Kebudayaan Berbasis Android," *Semin. Nas. Ris. dan Teknol. (SEMNAS RISTEK)*, vol. 4, no. 1, pp. 186–192, 2020.
[10] P. Nerurkar, A. Shirke, M. Chandane, and S. Bhirud, "Empirical Analysis of Data Clustering Algorithms," *Procedia Comput. Sci.*, vol. 125, pp. 770–779, 2018, doi: 10.1016/j.procs.2017.12.099.
[11] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Comput. Sci.*, vol. 148, pp. 291–302, 2019, doi: 10.1016/j.procs.2019.01.022.
[12] S. Reddy *et al.*, "Use and validation of text mining and cluster algorithms to derive insights from Corona Virus Disease-2019 (COVID-19) medical literature," *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. April, p. 100010, 2021, doi: 10.1016/j.cmpbup.2021.100010.
[13] S. Jun, S. S. Park, and D. S. Jang, "Document clustering method using dimension reduction and support vector clustering to overcome sparseness," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3204–3212, 2014, doi: 10.1016/j.eswa.2013.11.018.
[14] J. Rejito, A. Atthariq, and A. S. Abdullah, "Application of text mining employing k-means algorithms for clustering tweets of Tokopedia," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021, doi: 10.1088/1742-6596/1722/1/012019.
[15] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012016, 2021, doi: 10.1088/1742-6596/1783/1/012016.
[16] Mustakim *et al.*, "Unsupervised learning as a data sharing model in the fp-growth algorithm in determining the best transaction data pattern," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 11, pp. 2679–2689, 2021.
[17] M. T. Furqon and L. Muflikhah, "Clustering the potential risk of tsunami using Density-Based Spatial clustering of application with noise (DBSCAN)," *J. Environ. Eng. Sustain. Technol.*, vol. 3, no. 1, pp. 1–8, 2016.
[18] Mustakim, E. Rahmi, M. R. Mundzir, S. T. Rizaldi, Okfalisa, and I. Maita, "Comparison of DBSCAN and PCA-DBSCAN

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 884−895
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7512

Algorithm for Grouping Earthquake Area," in *2021 International Congress of Advanced Technology and Engineering, ICOTEN 2021*, 2021, pp. 0–4, doi: 10.1109/ICOTEN52080.2021.9493497.

[19] A. Fauzan, A. Novianti, R. R. M. A. Ramadhani, and M. A. S. Adhiwibawa, "Analysis of Hotels Spatial Clustering in Bali: Density-Based Spatial Clustering of Application Noise (DBSCAN) Algorithm Approach," *EKSAKTA J. Sci. Data Anal.*, pp. 25–38, 2022, doi: 10.20885/eksakta.vol3.iss1.art4.

[20] S. N. Yanti and E. Budiyati, "Aplikasi Pengenalan Budaya Provinsi Bagian Wita Di Indonesia Berbasis Android," pp. 1–11, 2020.

[21] S. S. Tafui, "Aplikasi Pengenalan Kebudayaan Kabupaten Belu Berbasi Android," *J. Mhs. Tek. Inform.*, vol. 1, no. 2, pp. 61–66, 2017.

[22] S. Hozeng and A. Syam, "Aplikasi Pengenalan Kebudayaan Khas Toraja (UKIRAN) Berbasis Android," *Semnasteknomedia*, vol. 5, no. 1, pp. 55–60, 2017.

[23] A. Atan, Z. Indra, and A. Febtriko, "Perancangan Game Berbasis Android Untuk Memperkenalkan Adat Melayu Riau," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 5, no. 1, pp. 54–66, 2020, doi: 10.36341/rabit.v5i1.963.

[24] N. Metafani and D. Djamaludin, "Aplikasi Pengenalan Cagar Budaya Tangerang Berbasis Android Di Dinas Kebudayaan Dan Pariwisata Kota Tangerang," *JIMTEK J. Ilm. Mhs. Fak. Tek.*, vol. 1, no. 1, pp. 66–73, 2020.

[25] Mustakim *et al.*, "DBSCAN algorithm: Twitter text clustering of trend topic pilkada pekanbaru," *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012001.

[26] K. N. S. Behara, A. Bhaskar, and E. Chung, "A DBSCAN-based framework to mine travel patterns from origin-destination matrices: Proof-of-concept on proxy static OD from Brisbane," *Transp. Res. Part C Emerg. Technol.*, vol. 131, no. August 2020, p. 103370, 2021, doi: 10.1016/j.trc.2021.103370.

[27] Q. Zhu, X. Tang, and A. Elahi, "Application of the novel harmony search optimization algorithm for DBSCAN clustering," *Expert Syst. Appl.*, vol. 178, no. April, p. 115054, 2021, doi: 10.1016/j.eswa.2021.115054.

[28] R. Novita, Mustakim, and F. N. Salisah, "Determination of the relationship pattern of association topic on Al-Qur'an using FP-Growth Algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012020, 2021, doi: 10.1088/1757-899x/1088/1/012020.

[29] L. Shabtay, P. Fournier-Viger, R. Yaari, and I. Dattner, "A guided FP-Growth algorithm for mining multitude-targeted item-sets and class association rules in imbalanced data," *Inf. Sci. (Ny).*, vol. 553, pp. 353–375, 2021, doi: https://doi.org/10.1016/j.ins.2020.10.020.

[30] J. Wang and Z. Cheng, "FP-Growth based Regular Behaviors Auditing in Electric Management Information System," *Procedia Comput. Sci.*, vol. 139, pp. 275–279, 2018, doi: 10.1016/j.procs.2018.10.268.