**Jurnal Teknik Informatika (JUTIF)** Vol. 6, No. 5, October 2025, Page. 3188-3203
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

# Sentiment Analysis and Topic Modeling for Discovering Knowledge in Indonesian Mobile Government Applications

**Ricky Bahari Hamid[1], Chandra Andriansyah[2], Dana Indra Sensuse[3], Sofian Lusa[4], Damayanti Elisabeth[5], Nadya Safitri[6]**

[1,2,3,4,5,6]Computer Science, Universitas Indonesia, Indonesia

Email: [1]ricky.bahari@ui.ac.id

Received : Jun 30, 2025; Revised : Aug 28, 2025; Accepted : Sep 2, 2025; Published : Oct 16, 2025

**Abstract**

The accelerated rate of government applications development in Indonesia has introduced new opportunities and challenges in delivering digital public services. While thousands of apps have been developed, systemic issues ranging from usability flaws to authentication failures persist, as reflected in user reviews on platforms like the Google Play Store. This study adopts a knowledge discovery approach to extract actionable insights from more than 17,000 user-generated reviews across three major government applications: Satusehat, Digital Korlantas, and M-Paspor. A hybrid methodology is applied, combining RoBERTa-based sentiment classification, BERTopic-based topic modeling, cosine similarity analysis, and qualitative user validation. The findings reveal recurring issues in authentication, interface design, and system responsiveness that span across organizational boundaries. Cross-app topic correlation highlights critical shared pain points such as login failures and unintuitive UI that undermine user trust in e-government services. Mapping these insights onto the SECI knowledge management model, this research contributes both practical recommendations and a replicable analytical framework for public agencies seeking to institutionalize user feedback. By transforming fragmented digital feedback into organizational knowledge, this study supports continuous service improvement and strengthens the foundation for user-centric e-government.

***Keywords :*** *BERTopic, e-Government, Knowledge Discovery , RoBERTa, Sentiment Analysis, Topic Modeling;*



## 1. INTRODUCTION

Governments worldwide are increasingly adopting digital technologies to deliver public services, where Indonesia is also undergoing this digital transformation process. This digital transformation aiming to enhance transparency, efficiency, and accessibility of services for the public [1]. In Indonesia, services based on government applications have expanded rapidly in recent years estimating the creation of more than 27000 apps consists of web application, desktop application and mobile application from both central government and local government [2] [3]. Along with the increasing number of government applications, the volume of user review left on various application distribution platform such as Google Playstore has also increased, which then formed as a rich data source, containing direct information about user satisfaction, complaints, expectation and suggestions [4]. Government failure to understand and responds to public feedback could result in negative impact leading to user frustration and reduced adoption rates [5].

A recent study found that applying sentiment analysis to social media can enhance transparency, encouraging citizen participation, and innovation in the government processes [6]. Internationally, similar approaches have been applied. For example, studies in China mined user-generated reviews of Government applications through sentiment analysis and topic modeling, revealing a critical service quality issues [7]. These findings underscore the potential of user feedback as a strategic input for continuous digital service improvement .

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

However, utilizing citizens feedback for continuous service improvement remains challenging as most prior research remains limited to single-application analysis, without exploring cross-application patterns that could reveal systemic issues or best practices [2].

Therefore, to address this gap this study will aim to do a knowledge discovery approach to uncover the common topic and patterns across multiple government apps by integrate sentiment analysis, topic modeling, and finding the topic correlation across those apps using user feedback as the dataset gathered from Google Play Store. Hence, it is hoped to bring novelty by proposing a cross-application approach that enables the identification of systemic issues in e-government public services. Specifically, it seeks to answer the following research questions (RQ):
RQ 1: What are the discovered topics or issues discussed by users in the reviews between application?
RQ 2: Are there any correlations between the identified topics across different government applications?
RQ 3: How can the insights derived from the topics be leveraged as knowledge to improve the e-government applications?

By answering these questions, this study makes two key contributions. Practically, it provides policymakers and developers with empirically grounded insights for enhancing public service applications. Theoretically, it advances knowledge discovery in e-government fields by offering a replicable analytical framework that transforms fragmented user feedback into structured, actionable knowledge.

## 2. METHOD

This research is divided into three phases, which is initiation, implementation, and evaluation and could be seen in figure 1. Initiation phase will focuses on data collection and data preprocessing. As for the implementation phase, the dataset that already went through the preprocessing process will be put into sentiment classification followed by the topic modeling. Lastly, the evaluation phase will be done with evaluating the sentiment analysis and topic modeling results, evaluate the correlation between discovered topics, and conduct user validation interview. These phases are sequentially designed to ensure systematic knowledge discovery from user-generated content, enabling actionable insights for e-government improvement.

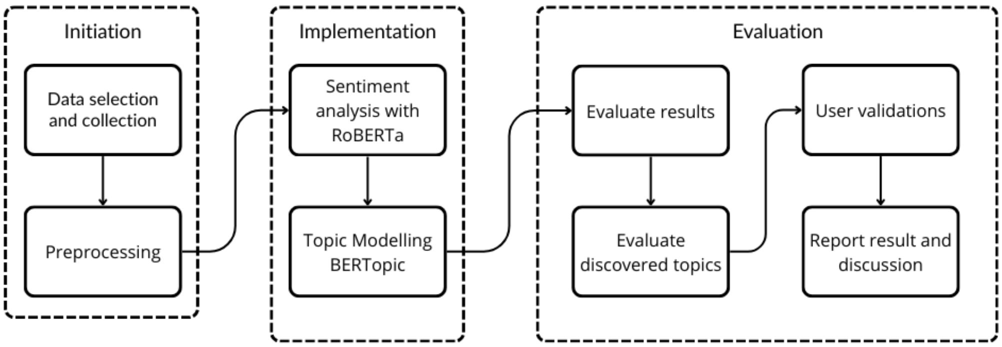

Figure 1. Methodology for research

The methodology is guided by prior findings in the domains of sentiment analysis, topic modeling, and knowledge discovery, each playing a crucial role in transforming user-generated app reviews into structured, actionable insights.

### 2.1. Related works

The methodology of this study is informed by key findings from previous research across multiple domains. Table 1 summarizes foundational studies that support the use of RoBERTa for sentiment

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

classification, BERTopic for topic modeling, and the integration of these tools into a knowledge management framework.

Table 1. Prior works summary

| No | Topics | Result |
|---|---|---|
| 1 | User-review mining for e-government [2] [8] [9] [10] [11] [12] | Establises that mining user reviews from application stores and social media, where its common in commercial software can likewise provide valuable, under-explored knowledge for improving Indonesian e-government apps. |
| 2 | Sentiment analysis with BERT [13] [14] [15] | Introduces sentiment analysis with IndoBERT and RoBERTa that show reliable sentiment classificaiton models for Indonesian language app reviews. |
| 3 | Topic modeling with BERTopic [16] [17] [18] [19] | Argues that topic modeling especially BERTopic reveals deeper factors behind user experience than sentiment alone, outperforming traditional LDA/LSA on short Indonesian texts. |
| 4 | Knowledge Discovery in KM [20] [21] | Connects topic modeling to the knowledge management concept of knowledge discovery, citing comparative studies that validate BERTopic as a powerful tool for extracting actionable insights from large text dataset. |

These prior works provide empirical justification for the tools and approaches selected in this study and highlight the relevance of user review mining for both technical development and knowledge generation in public sector contexts.

### 2.2. Data Selection and Data Collection

This knowledge discovery research hoped to yield a valuable knowledge especially for e-government advancement. Therefore, a set of inclusion rules was established for the selection of the mobile apps as shown in table 2.

Table 2. Application selection criteria

| ID | Inclusion Criteria |
|---|---|
| IC.1 | The app should deliver a core public service |
| IC.2 | The app should developed and operated by government |
| IC.3 | The app should have large user base |
| IC.4 | The app should have average rating < 4 |

After finding the application candidate, next was to collect the reviews by scraping the data from Google Play Store. Data collection performed using scraping techniques by utilizing Python library such as Google-Play-Scrapper. The collected dataset then filtered by the recent posted started from early 2025.

### 2.3. Data Preprocessing

Data preprocessing is a crucial step in preparing raw data for analysis in machine learning and NLP projects. RoBERTa and BERTopic were based on transformer models that is still requiring data preprocessing to achieve its optimal results such as data cleaning, case folding, normalization, stop word removal.

1. **Data cleaning**: Cleaning text data by eliminating irrelevant elements like punctuation, numbers, special characters, emojis, URLs, and meaningless words to enhance the accuracy and efficiency of text analysis

**Jurnal Teknik Informatika (JUTIF)**
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

2. **Case Folding**: Converting all text data to lowercase
3. **Normalization**: Because the dataset we collected are from an social media platform, most of it contains informal or slang words which is needed to be normalize first.
4. **Stop word removal**: Process of eliminating common words (such as "the," "is," "and") that do not carry significant meaning, in order to focus on more informative terms during text analysis.

### 2.4. Sentiment Analysis

In this research, we are going to do sentiment classification on the dataset to further enhance the topic modeling. Even though the user reviews on google play store have ratings, but it's not quite accurate. Figure 2 shows an example of a negative sentiment revies with relatively high rating.

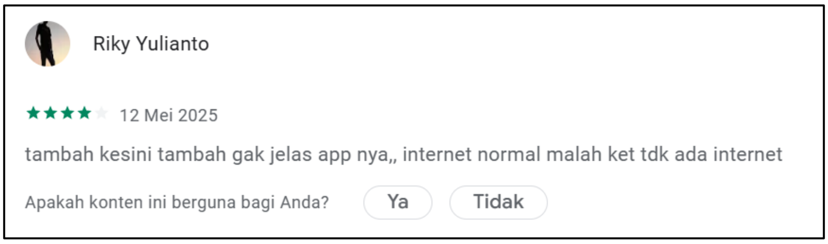


Figure 2. User review example

Therefore, sentiment classification were needed for the dataset to make sure sentiment of the reviews. In this research we are using Indonesian RoBERTa models, which is a sentiment-text-classification model based on RoBERTa model. It is a pre-trained model fine-tuned on a large dataset consisting of Indonesian comments and reviews [22].

### 2.5. Topic modeling

BERTopic is a modern topic modeling technique that leverages transformer-based language models such as BERT to generate dense semantic embeddings of textual data. Unlike traditional methods like Latent Dirichlet Allocation (LDA) that rely on word co-occurrence, BERTopic captures contextual meaning, making it highly effective for analyzing short and unstructured texts such as mobile application user reviews. This technique leverages transformer-based embeddings, reduces dimensionality using UMAP, and clusters data with HBDSCAN. Topics are then represented through class-based TF-IDF, enabling interpretable keywords groups, allowing for better coherence and relevance in the resulting topics [19]. Figure 3 shows the general flow of BERTopic technique.

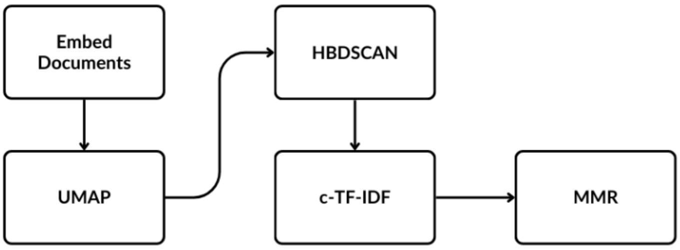


Figure 3. BERTopic modeling

### 2.6. Evaluation

Evaluation phase in this research mainly focuses on the topic created from the dataset, and the correlation between those topics for each application. To understand how users' expressed topics differ or align across various sentiments in mobile application reviews, this study implements cosine similarity

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

as an evaluation method. By comparing the semantic proximity of topic vectors within and across applications and sentiment classes (positive, neutral, negative), we aim to uncover patterns of user concern, satisfaction, and indifference. Cosine similarity is a widely used measure in various text analysis tasks, including topic correlation. It calculates the similarity between two vectors by measuring the cosine of the angle between them [23]. In topic modeling, cosine similarity can be used to measure the correlation between topics [24] [25], where it works on measuring the embedding vectors to see its semantic similarity. In cosine similarity each topic was represented as a vector in an n-dimensional space derived from BERTopic embeddings. The cosine similarity between topic vectors $T_i = (x_1, x_2, \dots, x_n)$ and $T_j = (y_1, y_2, \dots, y_n)$ with the formula as seen in (1).

$$\cos(\theta) = \frac{\sum_{i=1}^{n}(x_i y_i)}{\sqrt{\sum_{i=1}^{n}(x_i)^2}\sqrt{\sum_{i=1}^{n}(y_i)^2}} \qquad (1)$$

This approach allows us to assess not only the distinctiveness of topics across sentiment but also whether certain themes recur regardless of sentiment polarity. Cosine similarity is used to measure the similarity between popular topics.

## 3. RESULT

### 3.1. Data Selection and Data Collection

After all the selection criteria applied, three application candidate as seen in table 3 was selected which is Satusehat, M-paspor, and Digital Korlantas.

Table 3. Selected application information criteria

| Application | Description | Downloaded | Rating |
|---|---|---|---|
| Satusehat | An official health app by the Indonesian Ministry of Health that provides access to personal medical records, vaccination history, and integration with health facilities | 50+ million | 3.4 |
| Digital Korlantas | A traffic service app by the Indonesian National Police, offering online services such as driver's license (SIM) renewal, electronic traffic ticket (ETLE) info, and vehicle registration validation. | 10+ million | 2.9 |
| M-Paspor | An official app by the Directorate General of Immigration for passport applications and renewals, allowing users to fill out forms, upload documents, and schedule appointments at immigration offices. | 5+ million | 2.3 |

For the data collection, we are doing scraping of the google play store website to get all the needed user review data. Figure 4 shows scraping process using google-play-scraper library.

```
...  Memulai scraping untuk Google Play ID: com.telkom.tracencare
     Nama Aplikasi: SATUSEHAT Mobile
     Developer: Ministry of Health Republic of Indonesia
     Skor: 3.372648
     ------------------------------
```

Figure 4. Scrapping Process

The scraping process resulted in a total of 17362 data were collected from the scraping process, with detail as shown in table 4.

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

Table 4. Collected data

| Apps Name | Total Reviews |
|---|---|
| Digital korlantas | 7427 |
| Satusehat | 5550 |
| M-Paspor | 4385 |
| **Total** | **17362** |

### 3.2. Sentiment Analysis

In this phase, sentiment analysis was conducted on user reviews for the collected dataset using an Indonesian language transformer-based model which is indonesian RoBERTa. The objective of this analysis was to classify each review into one of three sentiment categories: Positive, Negative, or Neutral.

To evaluate the performance of the pre-trained RoBERTa model, a validation data consist of 20% of the total dataset  was manually labeled by two person. As for the evaluation metrics used include accuracy, precision, recall, and F1-score. Total validation data which totalled around 3350 data (20%) were a common ratio of testing or validation in sentiment analysis, were it is split evenly between sentiments, to ensure its accuracy [15] [26]. Table 5 shows the detail number of each sentiments whereas Figure 5 shows the spread to illustrate the balance of the validation dataset.

Table 5. Manual labeled dataset

| Manually Labeled | | |
|---|---|---|
| **Negative** | **Neutral** | **Positive** |
| 1196 | 1027 | 1127 |

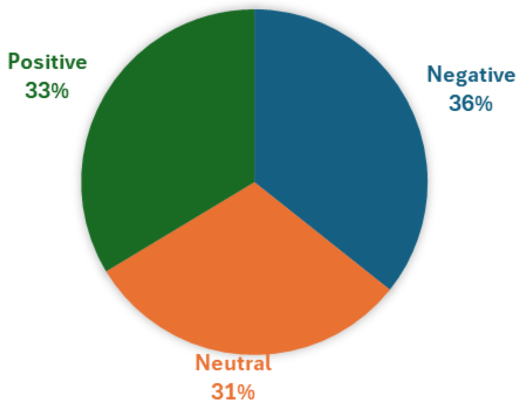


Figure 5. Spread of manual labeled data

The model's performance was assessed by comparing its sentiment predictions against the labels in the ground truth dataset. The precision, recall, and F1-score for each sentiment class Negative, Neutral, and Positive were consistently high. The Negative class achieved an F1-score of 0.95, while the Neutral and Positive classes achieved F1-scores of 0.93 and 0.97, respectively. These results reflect the model's balanced ability to correctly identify and classify various sentiment categories.. Detailed performance metrics are presented in Table 6.

**Jurnal Teknik Informatika (JUTIF)** Vol. 6, No. 5, October 2025, Page. 3188-3203
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

*Table 6. RoBERTa Performance Metric*

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| Negative | 0.98 | 0.92 | 0.95 |
| Neutral | 0.89 | 0.97 | 0.93 |
| Positive | 0.98 | 0.97 | 0.97 |
| **Overall Accuracy** | | | **0.95** |

For a more in-depth understanding of the model classification patterns and errors, the confusion matrix is presented in Figure 6. The confusion matrix confirms that the majority of predictions lie on the main diagonal, indicating correct classifications. The most significant error pattern is the misclassification of 88 Negative reviews as Neutral. This suggests that the model's greatest challenge lies in distinguishing between genuinely neutral expressions and those that are subtly or not explicitly negative.

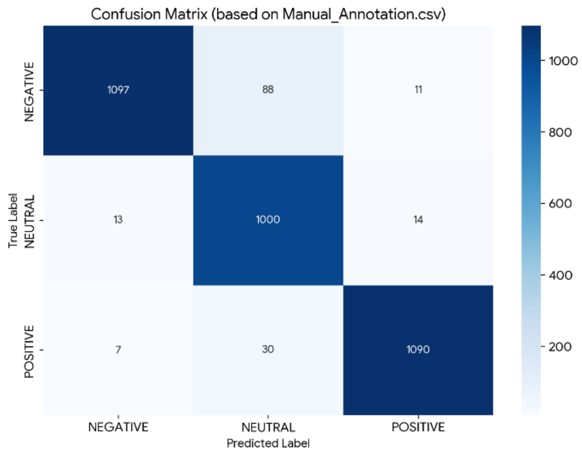


Figure 6. Confusion Matrix

With the high accuracy of the model, the classification of the rest of the dataset continue, with result shown in table 7.

*Table 7. Classification recap*

| Application Name | Negative | Neutral | Positive |
|---|---|---|---|
| Digital korlantas | 3601 | 926 | 2900 |
| M-paspor | 3427 | 402 | 556 |
| Satusehat | 3776 | 854 | 919 |
| **Total** | | **17361** | |

### 3.3. Topic Modeling with BERTopic

Following the sentiment labeling, BERTopic was used to uncover the latent topics of each application. For the clarity of this research, five most frequent topics per sentiment are presented for each application.

#### 3.3.1. Satusehat

Thirty distinct topics emerged in Satusehat results. Negative reviews are dominated by login loop/auto-logout (50%) and missing vaccine certificates (28%), pointing to authentication and data-retrieval failures. Other clusters include frequent errors, confusing UI, and NIK recognition problems.

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

Positive topics highlight smooth use cases such as "very helpful & easy" (33%) and "smooth vaccination process" (20%). Neutral reviews often describe the app as functional but slow (55%). Detailed topics are shown in Table 8.

*Table 8. Satusehat Topics*

| Sent. | Rank | ID | Topik | %dataset | Top-5 Keywords |
|---|---|---|---|---|---|
| **Negative** | 1 | 0 | Login loop / auto-logout | 50% | *email, login, akun, logout, kode* |
| | 2 | 1 | Missing vaccine certificate | 28% | *vaksin, sertifikat, download, hilang, booster* |
| | 3 | 2 | Frequent error | 15% | *aplikasi, pemerintah, ga, error, anggaran* |
| | 4 | 3 | Confusing UI layout | 3.70% | *ribet, ui, tampilan, gajelas, menu* |
| | 5 | 4 | NIK not recognised | 3.30% | *nik, daftar, invalid, data, sesuai* |
| **Positive** | 1 | 10 | Very helpful & easy | 33% | *mudah, membantu, cepat, praktis, mantap* |
| | 2 | 11 | Smooth vaccination process | 20% | *booster, registrasi, lancar, sertifikat, langsung* |
| | 3 | 12 | Lite-size & light | 17% | *ringan, cepat, size, ram, battery* |
| | 4 | 13 | Fast login | 16% | *login, cepat, sekali, klik, masuk* |
| | 5 | 14 | Helpful notifications | 14% | *notifikasi, reminder, vaksin, jadwal, push* |
| **Neutral** | 1 | 20 | App works but slow | 55% | *loading, lambat, tunggu, spinner, server* |
| | 2 | 21 | Data sometimes delayed | 22% | *data, update, telat, sertifikat, muncul* |
| | 3 | 22 | Update needed | 8% | *update, ui, versi, tolong, perbaiki* |

### 3.3.2. Digital Korlantas

Negative sentiment is led by SIM renewal failures (33%) and KTP/face-ID errors (13%), reflecting process fragility in core workflows. Other issues include app crashes, slow responses, and phone number rejections. Positive clusters include "online SIM extension works" (32%) and "very satisfied" (29%), validating the service's digital value when functioning correctly. Neutral reviews highlight reliability drops after updates ("used to be better," 28%). Topics are shown in Table 9.

Table 9. Digital Korlantas Topics

| Sent. | Rank | ID | Topik | %dataset | Top-5 Keywords |
|---|---|---|---|---|---|
| **Negative** | 1 | 0 | SIM renewal fails | 33% | *sim, perpanjang, gagal, proses, server* |
| | 2 | 1 | KTP / face-ID error | 13% | *ktp, wajah, verifikasi, gagal, scan* |
| | 3 | 2 | Frequent app crash | 9% | *force, close, crash, eror, aplikasi* |
| | 4 | 4 | Slow & no CS reply | 3% | *lambat, cs, balas, pelayanan, respon* |
| | 5 | 5 | Phone-number rejected | 3% | *nomor, hp, format, tidak, valid* |
| **Positive** | 1 | 10 | Online SIM extension works | 32% | *sim, online, berhasil, mudah, cepat* |
| | 2 | 11 | Very satisfied / mantap | 29% | *mantap, puas, bagus, terbaik, sukses* |
| | 3 | 12 | App is helpful | 16% | *aplikasi, membantu, praktis, urus, sim* |
| | 4 | 13 | Fast & easy process | 15% | *cepat, proses, mudah, selesai, upload* |
| | 5 | 14 | Service is fast | 8% | *pelayanan, cepat, satpas, waktu, singkat* |
| **Neutral** | 1 | 20 | Used to be better | 28% | *dulu, lancar, sekarang, berat, error* |
| | 2 | 21 | App & service OK | 22% | *aplikasi, oke, fungsi, lengkap, baik* |
| | 3 | 22 | New SIM option missing | 18% | *sim, baru, menu, belum, tersedia* |
| | 4 | 23 | Verification needs tweak | 17% | *verifikasi, ktp, ulang, gagal, perlu* |
| | 5 | 24 | Internet but still error | 15% | *sinyal, full, internet, error, jaringan* |

### 3.3.3. M-Paspor

In M-Paspor, negative sentiment is more fragmented, with multiple small clusters rather than one dominant issue. Key topics include "quota slot always full (14%)", maintenance downtime, OTP

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

failures, login problems, and frequent force-closes. This dispersion suggests that M_Paspor suffers from multi-point failures, requiring improvement in both infrastructure and service flow.

Positive reviews appear when the process runs smoothly ("app gets the job done," 29%) or improvements are noted in payment flow and appointment booking. Neutral reviews describe slow but functional use (67%). Topics are listed in Table 10.

Table 10 M-Paspor Topics

| Sent. | Rank | ID | Topik | Dataset% | Top-5 Keywords |
|---|---|---|---|---|---|
| **Negative** | 1 | 0 | Quota slot always full | 14% | *quota, slot, antrean, full, jadwal* |
| | 2 | 2 | Constant maintenance downtime | 6% | *maintenance, server, sibuk, kapan, lama* |
| | 3 | 5 | OTP never received | 4% | *otp, kode, email, sms, kirim* |
| | 4 | 13 | Cannot log in | 4% | *login, gagal, masuk, error, akun* |
| | 5 | 14 | App keeps force-closing | 4% | *crash, force, close, aplikasi, tiba* |
| **Positive** | 1 | 10 | App gets the job done | 29% | *berhasil, mudah, oke, selesai, paspor* |
| | 2 | 11 | Passport creation smooth | 21% | *paspor, buat, lancar, proses, cepat* |
| | 3 | 12 | Better payment flow | 18% | *pembayaran, virtual, account, berhasil, bank* |
| | 4 | 13 | Easier than before | 17% | *lebih, mudah, dulu, sekarang, simpel* |
| | 5 | 14 | Great despite crashes | 15% | *crash, kadang, tetap, mantap, fungsi* |
| **Neutral** | 1 | 20 | App opens but slow | 67% | *loading, lama, buka, tunggu, server* |
| | 2 | 21 | Pick appointment date | 16% | *tanggal, pilih, sulit, sesuai, jadwal* |
| | 3 | 22 | Want to make passport | 12% | *mau, paspor, baru, coba, nanti* |
| | 4 | 23 | Office visit still needed | 5% | *kantor, imigrasi, harus, foto, datang* |
| | 5 | 24 | Internet but still error | 15% | *sinyal, full, internet, error, jaringan* |

### 3.3.4. Word Cloud Visualization

Figure 8 presents a word cloud summarizing frequent keywords from user reviews across the three applications. In M-Paspor, negative terms such as error, lama, and lambat highlight delays and instability, while neutral terms like paspor and imigrasi describe routine transactions. In Satusehat and Digital Korlantas, positive words such as "cepat", "baik", and "membantu" reflect satisfaction with speed and usability, whereas negative words like "ribet" indicate confusion with updates and interface design. Overall, the visualization reinforces the recurring themes identified in topic modeling.


*Figure 7. Word Cloud*

## 3.4. Evaluation

### 3.4.1. Qualitative Validation

Nine semi-structured interviews (three users per application) were coded line-by-line against the top five negative topics for each app. These interviews were intended as a means of validations in terms

**Jurnal Teknik Informatika (JUTIF)** Vol. 6, No. 5, October 2025, Page. 3188-3203
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

of biases among users. Table 11-13 records topic presence a check mark indicates the respondent experienced the issue exactly as phrased in the interview question.

Table 11. Satusehat user validation

| App / Respondent | Login loop | Missing cert | Frequent error | Confusing UI | NIK error |
|---|---|---|---|---|---|
| RS1 | | | | ✓ | ✓ |
| RS2 | ✓ | | | ✓ | |
| RS3 | | ✓ | | | |

Table 11 shows that two out of three users (66 %) confirmed the Confusing UI issue while one also experienced the Login problems. These responses closely mirror the top negative clusters identified in the topic modeling results.

Table 12. Digital Korlantas user validation

| App / Respondent | SIM renewal fails | KTP / Face-ID error | App crash | Slow / no CS | Phone # rejected |
|---|---|---|---|---|---|
| RSD1 | | | | | |
| RSD2 | ✓ | ✓ | ✓ | | |
| RSD3 | ✓ | ✓ | | ✓ | |

In table 12, for Digital Korlantas, both SIM-renewal failures and KTP/Face-ID errors were reported by two of the three participants (66%). This aligns precisely with the two largest negative clusters, confirming that authentication and workflow reliability remain the app's most critical weaknesses.

Table 13. M-Paspor user validation

| App / Respondent | Quota slot full | Maintenance downtime | OTP never received | Cannot log in | Force-close |
|---|---|---|---|---|---|
| RSM1 | ✓ | | | | |
| RSM2 | ✓ | | | | |
| RSM3 | ✓ | ✓ | | ✓ | ✓ |

As for M-Paspor user validation shown in table 13, All respondents (100 %) encountered "Quota slot always full", the model's highest-share topic, while at least one user validated each of the next three clusters (Maintenance downtime, Cannot log in, Force-close).

Overall, these validations confirm that the clusters produced by BERTopic accurately reflect user experiences, strengthening confidence in the results.

#### 3.4.2. Evaluation of Topic Correlation across Application

To understand how users' expressed topics differ or align across various sentiments in mobile application reviews, this study implements cosine similarity as an evaluation method. By comparing the semantic proximity of topic vectors within and across applications and sentiment classes (positive, neutral, negative), we aim to uncover patterns of user concern, satisfaction, and indifference. This approach allows us to assess not only the distinctiveness of topics across sentiment but also whether certain themes recur regardless of sentiment polarity.

**Jurnal Teknik Informatika (JUTIF)**
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

We adopt a similar strategy tailored to mobile app reviews where three cosine-similarity matrices were prepared between Satusehat, Digital Korlantas, and M-Paspor, each comparing every pair of negative-sentiment topics. Figures 9-10 visualise the matrices as heat-maps, while table 14 shows the topic correlation with one or both topic included in top five topics. A topic pair is deemed correlated when its cosine value is ≥ 0.80, and lower values are considered mere semantic neighbours.

Table 14. Cross-application topic correlations with cosine similarity ≥ 0.80

| Matrix | Topic ID | Cos sim | Common Theme |
|---|---|---|---|
| M-Paspor – Satusehat | **T 15 - T 3** | **0.98** | Bad UI |
| M-Paspor - Digital Korlantas | **T 13 - T 7** | **0.94** | Login / authentication failure |
| Satusehat - Digital Korlantas | **T 0 - T 2** | **0.92** | Session break / repeated login |
| M-Paspor - Digital Korlantas | **T 3 - T 2** | **0.92** | Error opening the App |
| Satusehat - Digital Korlantas | **T 9 - T 4** | **0.88** | Slow response & latency |

The cross-matrix inspection reveals three clear patterns. First, authentication emerges as a ministry-wide weak-spot. The closest functional match (cos = 0.94) links M-Paspor's "Cannot-login/OTP delay" (T 13) with Digital Korlantas' "Login gagal–susah" (T 7), while Satusehat "Login loop" (T 0) aligns strongly (cos = 0.92) with Korlantas' "Session break" (T 2). These pairings show that identity-verification failures transcend sector boundaries.

A second prominent pattern involves interface design and performance stability. The highest overall score (cos = 0.98) merges M-Paspor's *"jelek–ribet UI" (T15)* with Satusehat's *"ribet–gajelas UI" (T3)*, reflecting a shared perception of unintuitive government app interfaces. Similarly, Satusehat's *"Very slow" (T9)* correlates with Korlantas' *"Lambat" (T4)* at (cos = 0.88), highlighting widespread performance instability.

Third, certain issues remain domain-specific. M-Paspor's *quota-slot exhaustion* and Korlantas' *SIM-renewal workflow failures* have no counterparts above the 0.80 threshold, indicating these problems are tied to agency-specific processes rather than systemic weaknesses.

Overall, these high-similarity pairs supply quantitative proof that negative topics correlate across different government applications, thereby resolving RQ 2. Two sets of actions follow. Figures 9–11 illustrate these correlations, where yellow cells (cos ≥ 0.80) denote strong semantic overlap and blue cells (≤ 0.20) indicate no meaningful relation.

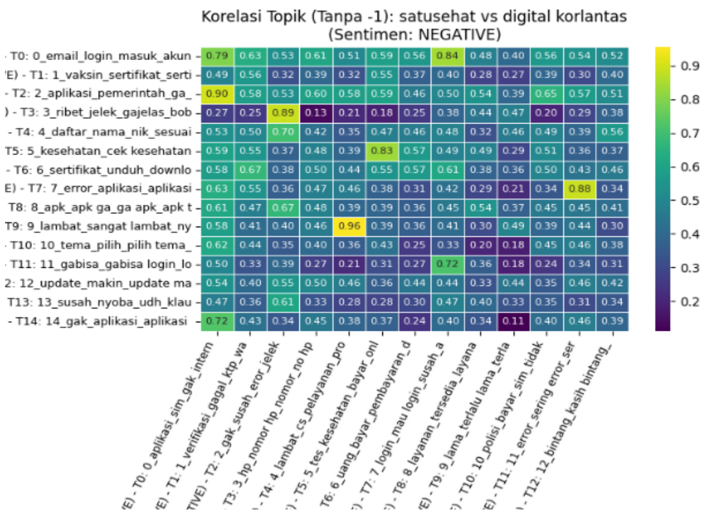


*Figure 8. Topic correlation heatmap: Satusehat vs Digital Korlantas*

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

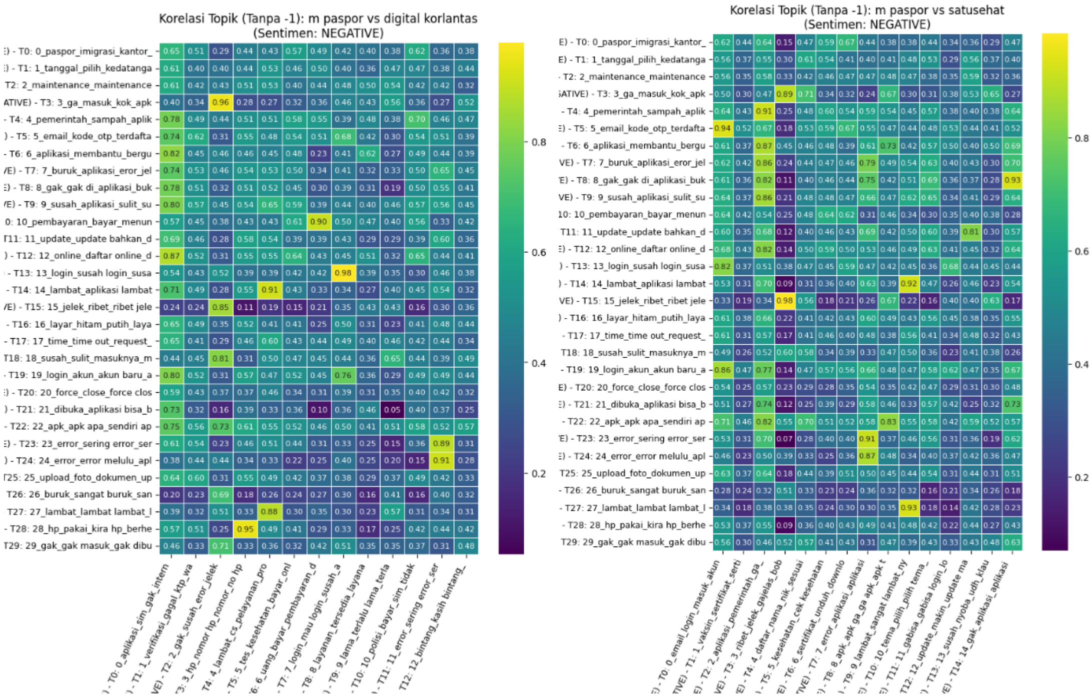


*Figure 9. Topic correlation heatmap: M-Paspor vs Digital Korlantas and M-Paspor vs Satusehat*

## 4. DISCUSSIONS

### 4.1. Synthesis of Cross-Application Findings and Their Implications

The primary strength of this research lies in its cross-application methodology, which was applied to government applications with large user bases: Satusehat, Digital Korlantas, and M-Paspor. By not only analyzing each application in isolation but also cross-referencing them using sentiment analysis, topic modeling, and cosine similarity correlation, it reveals systemic issues that cut across agencies rather than isolating problems within a single app.

Findings highlight three recurring weaknesses which is Authentication failures involving logins, OTPs, and identity verification. Second is interface and performance issues such as confusing UI/UX, latency, and slow responsiveness, and lastly about system instability, including frequent crashes and force-closes. These patterns confirm that user complaints reflect fundamental weaknesses in Indonesia's digital service ecosystem.

These correlated findings confirm that the problems users face are not isolated incidents but are reflections of fundamental weaknesses within the government's digital service ecosystem. These findings are consistent with global studies, for example the European mGov4EU project evaluated mobile government services and showed that adherence to usability Good Practices such as minimalistic design, error handling, and accessibility correlates strongly with improved user satisfaction and trust [27]. Similarly, a comparative study of Australian government mobile applications applied sentiment and topic modeling across services and demonstrated that cross-app feedback analysis is an effective way to identify systemic weaknesses and align services with digital transformation goals [28]. This emphasize that the systemic weaknesses identified in Indonesian apps mirror global challenges, while the novelty of this study lies in applying a cross-application approach in the Indonesian context.

This research presents not only a diagnosis of the problems but also introduces a replicable workflow mapped to the SECI knowledge management framework (Table 15), showing how

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

unstructured user feedback can be transformed into structured, actionable knowledge. This continuous improvement cycle can be adopted across agencies to systematically enhance digital service quality.

### 4.2. Implication on Knowledge Management

This research systematically transformed the insights derived from user reviews of government mobile applications into actionable knowledge assets. Following the SECI (Socialization, Externalization, Combination, Internalization) knowledge management framework, we explicitly connected user-driven insights to practical action as shown in table 15.

*Tabel 15. SECI Phase*

| SECI phase | Concrete action in this research | Output |
|---|---|---|
| Externalization (Tacit > Explicit) | 1. Collect raw user feedback<br>2. Classify sentiment using Indonesian RoBERTa<br>3. Extract stuctured topic labels using BERTopic clustering | Structured topic clusters with explicit labels and sentiment labels |
| Combination (Explicit > Explicit) | Measure cosine similarity matrices to identify correlated issues across different apps | Explicit cross-app correlation table (cosine similarity matrix) |
| Internalization (Explicit > Tacit) | Integrate explicit findings into organizational processes | Apply actionable recommendations from findings |

Findings from this KM Cycle were translated into actionable recommendations, hoped to bring benefits to both governments and the citizens that are using the public service applications. Those recommendations are:

1. High correlations in topics regarding authentication problems indicating that there are needs to standardize and enhance the authentication system in the public services apps. The concrete recommendations are to create cross-organization ad-hoc team to design and adopt a robust Single Sign-On solution (SSO) that reliable and safe.
2. Issues about interfaces also arises in topic correlations, suggesting the needs to develop a national UI/UX guidelines for government applications. Besides that, a joint training for all the government developers and regulations about usability testing before launching an applications also crucial to answer this issues.
3. Another high correlation topic is talking about the latency and slow response of the application. Therefore the actionable recommendation for this issue is to create guidelines for optimalization of IT infrastructure and application development, where use of relevant technologies such as cloud computing, CDN, and Clean Code practices to ensuring the quality of the application.

Comparable approaches are also evident internationally. For example, the Europe mGov4EU project emphasized how design Good Practices can be institutionalized into governance frameworks, effectively transforming usability insights into knowledge assets that guide future development [27]. What distinguishes this study, however, is the integration of multiple analytical layers into a unified SECI-based cycle. By applying sentiment classification, topic modeling, topic correlation, and qualitative validation together, this study ensures that user-generated feedback is not only interpreted but institutionalized as a structured knowledge resource. This provides a replicable framework for governments both in Indonesia and globally seeking to embed citizen input into digital transformation strategies.

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

## 5. CONCLUSION

This study examined user-generated Google Play reviews of three Indonesian government mobile applications Satusehat, Digital Korlantas, and M-Paspor using a mixed-methods pipeline that combined RoBERTa-based sentiment classification, BERTopic clustering, cosine-similarity analysis, and semi-structured interviews. The primary objectives were to identify dominant negative topics within each application (RQ 1), determine whether any topics overlapped across different apps (RQ 2), and convert these insights into actionable knowledge assets to improve e-government services (RQ 3).

Findings showed that three categories of problems dominated user dissatisfaction: authentication failures (login loops, OTP delays, KTP/face-ID errors), UI/UX and performance fragility (confusing interfaces, crashes, latency), and capacity bottlenecks (e.g., quota-slot exhaustion, SIM-renewal workflow failures). Cross-application analysis confirmed that authentication and UI/performance problems are systemic, while bottlenecks are app-specific. Semi-structured interviews with nine end users confirmed confirmed the reliability of these patterns as genuine user experiences.

The urgency of this research lies in demonstrating that these weaknesses are not isolated technical bugs but systemic flaws that threaten citizen trust in e-government services. With more than 27,000 government applications already deployed in Indonesia, recurring issues driving citizens back to offline services, undermining adoption, wasting resources, and slowing the national digital transformation agenda. By providing the cross-application evidence in the Indonesian context, this study fills a critical policy gap, offering stakeholders the knowledge base to justify cross-agency reforms such as single sign-on (SSO), standardized UI/UX guidelines, and infrastructure optimization.

This study contributes by mapped the entire pipeline onto the SECI (socialization, externalization, combination, internalization) model of knowledge creation. Of all the phase, Socialization were excluded because there are no direct interaction in the first phase of the study, instead this study are doing the interview to validate the found topics. This newfound knowledge then analyzed and become an actionable recommendation for the government to improve its e-government implementation, and gained public trust. This offers a replicable blueprint not only for Indonesian agencies but also for other governments.

Despite these contributions, several limitations warrant consideration. The analysis relied solely on Google Play reviews, which may not reflect experiences on iOS or alternative platforms. The qualitative sample ($n = 9$) was sufficient for initial validation but may not capture all demographic or contextual variations. Future research should expand interview sampling, include cross-platform data sources, and conduct longitudinal evaluations of user sentiment following the deployment of the proposed knowledge assets.

In conclusion, this research advances both e-government practice and knowledge-management theory by demonstrating a knowledge discovery in database (KDD) process from how user-generated feedback can be systematically mined, validated, transformed until become an actionable organizational knowledge. The resulting framework and artefacts provide a replicable blueprint for other public agencies to leverage citizen input, improve digital services, and foster continuous organizational learning.

## CONFLICT OF INTEREST

The authors declares that there is no conflict of interest between the authors or with research object in this paper.

## ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Ministry of Communication and Digital (Komdigi) for providing the funding support required for the submission of this paper. Appreciation is also extended to all individuals and organizations who contributed to the successful completion of this research.

Vol. 6, No. 5, October 2025, Page. 3188-3203
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.5.4991

## REFERENCES


[1] N. Aleisa, "Key factors influencing the e-government adoption: a systematic literature review," *Journal of Innovative Digital Transformation*, vol. 1, no. 1, pp. 14–31, Aug. 2024, doi: 10.1108/jidt-09-2023-0016.

[2] I. Nurkumalawati and M. S. Rofii, "Public Review of M-Paspor Application in Indonesia: Mobile Government, Digital Resilience, Cyber Security," in *Proceedings of the Fifth Annual International Conference on Business and Public Administration (AICoBPA 2022)*, 2023, pp. 404–412. doi: 10.2991/978-2-38476-090-9_32.

[3] Humas Menpan RB, "Dorong Efisiensi dan Keterpaduan Aplikasi dengan Proses Clearance." Accessed: Aug. 25, 2025. [Online]. Available: https://www.menpan.go.id/site/berita-terkini/dorong-efisiensi-dan-keterpaduan-aplikasi-dengan-proses-clearance?utm_source=chatgpt.com

[4] A. Ohoilulin, D. I. Inan, R. Juita, and M. Sanglise, "Dissatisfaction of a Mobile-Based Application from Different Platforms Using Naïve Bayes for Sentiment Analysis and LDA for Topic Modelling," *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 2, pp. 991–1001, Sep. 2024, doi: 10.47065/bits.v6i2.5729.

[5] A. K. Abdulkareem and R. Mohd Ramli, "Does trust in e-government influence the performance of e-government? An integration of information system success model and public value theory," *Transforming Government: People, Process and Policy*, vol. 16, no. 1, pp. 1–17, Feb. 2022, doi: 10.1108/TG-01-2021-0001.

[6] S. Verma, "Sentiment analysis of public services for smart society: Literature review and future research directions," *Gov Inf Q*, vol. 39, no. 3, p. 101708, Jul. 2022, doi: 10.1016/j.giq.2022.101708.

[7] N. Xu and W. Zhang, "User Satisfaction with Chinese Government Apps: Topic Mining and Sentiment Analysis of User Reviews," *Lex localis - Journal of Local Self-Government*, pp. 95–124, Jul. 2025, doi: 10.52152/23.3.95-124(2025).

[8] T. Liu *et al.*, "ROSEMATCHER: Identifying the impact of user reviews on app updates," *Inf Softw Technol*, vol. 161, Sep. 2023, doi: 10.1016/j.infsof.2023.107261.

[9] A. Purnama, M. L. Fitria, R. Y. Widyaputra, and Sunardi, "User Experience Evaluation on Mobile Application for Indonesian Passport Using Sentiment Analysis and Cognitive Walkthrough," in *Proceedings of 2024 International Conference on Information Management and Technology, ICIMTech 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 753–758. doi: 10.1109/ICIMTech63123.2024.10780800.

[10] B. Chow and N. Legowo, "Factors Influencing User Satisfaction of PeduliLindungi App with UTAUT & Delone Mclean Models: A Case Study in Indonesia," *Journal of System and Management Sciences*, vol. 13, no. 2, pp. 170–184, 2023, doi: 10.33168/JSMS.2023.0212.

[11] S. Malodia, A. Dhir, M. Mishra, and Z. A. Bhatti, "Future of e-Government: An integrated conceptual framework," *Technol Forecast Soc Change*, vol. 173, p. 121102, Dec. 2021, doi: 10.1016/j.techfore.2021.121102.

[12] X. Ye, X. Su, Z. Yao, L. A. Dong, Q. Lin, and S. Yu, "How Do Citizens View Digital Government Services? Study on Digital Government Service Quality Based on Citizen Feedback," *Mathematics*, vol. 11, no. 14, Jul. 2023, doi: 10.3390/math11143122.

[13] M. Wankhade, A. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif Intell Rev*, vol. 55, pp. 5731–5780, 2022, doi: 10.1007/s10462-022-10144-1.

[14] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, p. 2023, 2023, doi: 10.14569/IJACSA.2023.0140813.

[15] A. P. Rifai *et al.*, "DETECTION MODEL FOR FAKE NEWS ON COVID-19 IN INDONESIA," *ASEAN Engineering Journal*, vol. 13, no. 4, pp. 119–126, 2023, doi: 10.11113/aej.V13.19648.

[16] O. Grljević, M. Marić, and R. Božić, "Exploring Mobile Application User Experience Through Topic Modeling," *Sustainability*, vol. 17, no. 3, p. 1109, Jan. 2025, doi: 10.3390/su17031109.

[17] M. Mendonça and Á. Figueira, "Topic Extraction: BERTopic's Insight into the 117th Congress's Twitterverse," *Informatics*, vol. 11, no. 1, Mar. 2024, doi: 10.3390/informatics11010008.

[18] M. Ananta, R. Utari, A. Akhyar, and G. A. Putri Saptawati, "Integrating BERTopic and Large Language Models for Thematic Identification of Indonesian Legal Documents," in *2024 11th International Conference on Advanced Informatics: Concept, Theory and Application, ICAICTA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICAICTA63815.2024.10762984.

[19] M. F. Asnawi, M. Hanafi, N. F. Kurniawan, A. Suwondo, A. Nasrullah, and C. Setyawan, "Topic Modelling Analysis on Indonesian News Using BERT Topic Model," in *2024 6th International Conference on Cybernetics and Intelligent System, ICORIS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICORIS63540.2024.10903779.

[20] Becerra-Fernandez, Irma, Sabherwal, and Rajiv, "Knowledge Management Systems and Processes Second Edition," 2015.

[21] M. Hankar, M. Kasri, and A. Beni-Hssane, "A comprehensive overview of topic modeling: Techniques, applications and challenges," May 01, 2025, *Elsevier B.V.* doi: 10.1016/j.neucom.2025.129638.

[22] W. Wongso, "indonesian-roberta-base-sentiment-classifier (Revision ac452dc)," 2024, *Hugging Face*. doi: 10.57967/hf/2997.

[23] B. Li and L. Han, "Distance Weighted Cosine Similarity Measure for Text Classification," 2013, pp. 611–618. doi: 10.1007/978-3-642-41278-3_74.

[24] R. K. Roul and J. K. Sahoo, "Document Labeling Using Source-LDA Combined with Correlation Matrix," 2019, pp. 697–707. doi: 10.1007/978-981-10-8055-5_62.

[25] C. Lu, L. Zhu, Y. Xie, W. Xu, Y. Zhao, and Y. Cao, "Analysis of Hot Topics and Evolution of Research in World-class Agricultural Universities Based on BERTopic," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/amns-2024-0327.

[26] S. A. Ajagbe, J. B. Awotunde, and H. Florez, "Intrusion Detection: A Comparison Study of Machine Learning Models Using Unbalanced Dataset," *SN Comput Sci*, vol. 5, no. 8, 2024, doi: 10.1007/s42979-024-03369-0.

[27] R. Sellung and L. Kiss, "Evaluating the Good Practices of User Experience for Mobile and Electronic Governmental Services," in *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, Gesellschaft fur Informatik (GI), 2024, pp. 39–50. doi: 10.18420/OID2024_03.

[28] T. Aamir, M. B. Chhetri, M. A. P. Chamikara, and M. Grobler, "Government Mobile Apps: Analysing Citizen Feedback via App Reviews," in *Proceedings - 2023 38th IEEE/ACM International Conference on Automated Software Engineering, ASE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 1858–1863. doi: 10.1109/ASE56229.2023.00190.