

Implementation of SVM and GLDA for SERVQUAL Based-Gap Analysis on Mobile JKN Reviews

Balqis Nur Aura Shaviradilla^{1*}, I Kadek Dwi Nuryana¹ , Mochamad Nizar Palefi Ma'ady² 

^{1*}Universitas Negeri Surabaya, Surabaya, Indonesia

balqis.21028@mhs.unesa.ac.id, dwinuryana@unesa.ac.id

²Telkom University, Surabaya, Indonesia

mnizarpm@telkomuniversity.ac.id

Abstract. This study aims to analyze user perceptions of the Mobile JKN application developed by BPJS Kesehatan by integrating sentiment analysis, topic modeling, and service quality gap analysis. User reviews were collected from the Google Play Store using a scraping technique, resulting in 44,752 reviews. After removing duplicate entries, 30,926 reviews were retained, and 6,186 reviews were used for modeling and analysis. Sentiment classification was conducted using the Support Vector Machine (SVM) algorithm with linear and Radial Basis Function (RBF) kernels. The results show that the majority of user reviews contain negative sentiment, with 2,784 negative reviews identified using the linear kernel and 2,847 negative reviews using the RBF kernel. Topic modeling was performed using Guided Latent Dirichlet Allocation (GLDA) to map review topics into the five SERVQUAL dimensions: tangibles, assurance, empathy, reliability, and responsiveness. The GLDA model achieved a coherence score of 0.64 and a UMass score of -1.99, indicating that the generated topics are sufficiently interpretable and consistent. Gap analysis was then calculated by comparing the proportion of perceived performance derived from sentiment classification with the expected service representation obtained from topic distribution. The results indicate that the reliability dimension has the largest gap (-0.74), while assurance has the smallest gap (-0.06). The overall SERVQUAL gap value is 0.11 (11%), indicating that user satisfaction with the Mobile JKN application remains relatively low and that several service dimensions still require improvement.

Keywords: *Sentiment Analysis; Guided Latent Dirichlet Allocation; SERVQUAL; Support Vector Machine; Mobile JKN*

1. Introduction

Technological developments have significantly transformed healthcare services. Digital health innovation enhances service quality and public well-being through mobile applications (Giebel et al., 2024). One such app in Indonesia is Mobile JKN, developed by BPJS Kesehatan to help users access features like participant status, nearby hospitals, and treatment history (<https://jknmobile.com>). Despite a 4.3 rating and over 630,000 reviews on the Google Play Store, many five-star ratings include negative feedback, questioning the reliability of ratings. Since

90% of users check reviews before downloading apps, it means understanding user sentiment is critical.

Previous studies have applied machine learning techniques to analyze user feedback from large volumes of textual data. Support Vector Machine (SVM) has been widely used for sentiment classification due to its strong performance in text-based datasets ([Haoues et al., 2023](#)), ([Rahardi et al., n.d.](#)). In addition, topic modeling methods such as Latent Dirichlet Allocation (LDA) are commonly employed to identify dominant themes within user reviews ([Venugopalan & Gupta, 2022](#)). Meanwhile, service quality evaluation is frequently conducted using the SERVQUAL framework to measure gaps between perceived service performance and user expectations ([Kermanshachi et al., 2022](#)).

However, many studies evaluating service quality rely primarily on survey-based SERVQUAL methods with relatively limited respondent samples ([Sharifi et al., 2021](#)). Although these approaches provide structured evaluations, they may suffer from response bias and limited representativeness. On the other hand, sentiment analysis studies on public service applications often focus mainly on classification performance without linking the results to service quality dimensions ([Ossai & Wickramasinghe, 2023](#)). Similarly, topic modeling studies generally extract themes from user reviews but rarely connect the findings to structured service quality frameworks.

Recent research has also explored different methodological approaches for sentiment analysis. Traditional machine learning algorithms such as Support Vector Machine (SVM) remain widely used due to their effectiveness in handling high-dimensional textual features such as TF-IDF. In contrast, deep learning approaches such as Long Short-Term Memory (LSTM) and transformer-based models like BERT are increasingly applied because they can capture contextual semantic relationships within text and often achieve higher accuracy. However, these models generally require larger datasets and higher computational resources compared to traditional machine learning methods ([Sanwal & Muhammad, 2023](#)).

In topic modeling, Latent Dirichlet Allocation (LDA) is one of the most widely used probabilistic models for discovering latent topics within large text collections. Nevertheless, topics generated by standard LDA are sometimes difficult to interpret because the model relies entirely on unsupervised learning. To address this limitation, Guided Latent Dirichlet Allocation (Guided LDA) extends the LDA framework by incorporating predefined seed words to guide the topic generation process. This approach allows the discovered topics to align with predefined analytical frameworks, such as service quality dimensions ([Li et al., 2023](#)).

Despite these developments, studies integrating sentiment analysis, topic modeling, and structured service quality evaluation using large-scale user review data remain limited. Therefore, this study addresses this research gap by integrating Support Vector Machine (SVM) for sentiment classification, Guided Latent Dirichlet Allocation (GLDA) for topic identification, and SERVQUAL-based gap analysis within the CRISP-DM framework to analyze user reviews of the Mobile JKN application.

This study aims to analyze user perceptions of the Mobile JKN application by integrating sentiment analysis, topic modeling, and service quality evaluation. Specifically, the objectives of this study are to analyze user sentiment toward the Mobile JKN application using the Support Vector Machine (SVM) method, to identify the main topics discussed in Mobile JKN user reviews using the Guided Latent Dirichlet Allocation (GLDA) algorithm, and to evaluate the service quality gap based on the results of sentiment analysis and topic modeling using SVM and GLDA.

The significance of this study can be viewed from several perspectives. From a knowledge advancement perspective, this study proposes an integrated analytical framework that combines sentiment analysis using SVM, topic modeling using Guided LDA, and SERVQUAL-based gap analysis to evaluate user perceptions of mobile health service applications. From a practical

perspective, the findings provide insights for BPJS Kesehatan developers to identify service dimensions that require improvement, particularly reliability and responsiveness. From a societal perspective, the study supports the improvement of digital public service applications by utilizing user-generated reviews as a source of evidence-based service evaluation.

2. Methods

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which consists of several phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The research flow is illustrated in Figure 1.

2.1. Business Understanding

The purpose of this research is to: (1) conduct sentiment analysis using the Support Vector Machine (SVM) algorithm on user reviews of the Mobile JKN application; (2) conduct topic modeling using Guided Latent Dirichlet Allocation (GLDA); and (3) perform gap analysis using the results from SVM and GLDA.

2.2. Data Understanding

At the data understanding stage, exploratory data analysis (EDA) was conducted to obtain an overview of the dataset. This analysis includes examining the distribution of user ratings, identifying the frequency of reviews over time, and observing common terms appearing in the review texts. The EDA process helps identify potential patterns and data characteristics before performing further text preprocessing and modeling.

User reviews were collected from the Google Play Store between 1 January and 5 October 2024, resulting in 44,752 reviews. After removing duplicate entries, 30,926 reviews remained. From this dataset, 6,186 reviews were used for sentiment classification and topic modeling after preprocessing and data preparation.

The distribution of user ratings indicates that the Mobile JKN application generally receives high ratings on the Google Play Store, with an average rating of 4.3. However, several reviews with high ratings still contain negative feedback, suggesting inconsistencies between numerical ratings and textual opinions. This phenomenon highlights the importance of analyzing textual reviews rather than relying solely on rating values. Figure I shows review with high ratings still contain negative feedback, suggesting inconsistencies between numerical ratings and textual opinions.

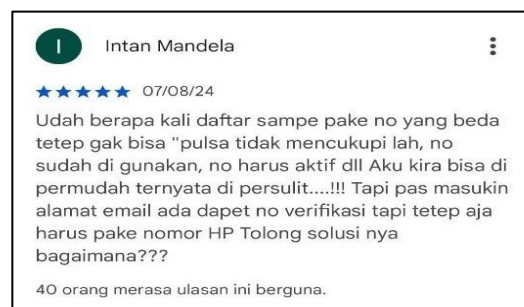


Figure 1 Mobile JKN User Review

2.3. Data Preparation

The raw review data underwent a series of preprocessing steps:

- Cleaning: Removing numbers, symbols, whitespace, punctuation, and emoticons.
- Case Folding: Converting all text to lowercase.

- Normalization: Using a standard dictionary for synonym replacement.
- Tokenization and Stopword Removal: Splitting text into words and eliminating irrelevant words.

From 44,752 scraped reviews, 30,926 reviews remained after cleaning and duplicate removal. Due to class imbalance and computational efficiency considerations, 6,186 reviews were selected using stratified sampling for model training and evaluation.

Labeling was conducted using the InSet Lexicon to classify each review into sentiment categories ([Barik & Misra, 2024](#)).

TF-IDF transformation was applied before feeding data into the model ([Alemerien et al., 2024](#)). Three train-test split ratios were tested: 80:20, 70:30, and 50:50 ([Bichri et al., 2024](#)). The combination of kernel type and data split ratio with the best accuracy was selected for the final analysis.

2.4. Modeling

2.4.1 Sentiment Analysis using SVM

SVM was applied using both linear and Radial Basis Function (RBF) kernels, and the kernel with the highest accuracy was selected. Sentiment categories were mapped to Likert scale values, as shown in Table 1.

Table 1 Classification of Sentiment

Scale	Sentiment
1	Not Satisfied
2	Moderately Satisfied
3	Satisfied

2.4.2 Topic Modeling Using GLDA

GLDA was used to generate topics based on SERVQUAL dimensions: Tangibles, Reliability, Responsiveness, Assurance, and Empathy ([Qiu et al., 2024](#)), ([Alanzi, 2022](#)), ([Zaim Mohd Salleh et al., 2024](#)). Seed words were defined and transformed into seed texts. Reviews were vectorized using CountVectorizer and input into the GLDA model, which was trained with predefined topics, iterations, and seed confidence. The model was evaluated using Coherence Score (c_v) and UMass ([Venugopalan & Gupta, 2022](#)).

2.4.3 Mapping Topics to SERVQUAL Dimensions

Topics generated from GLDA were decoded and filtered to map the reviews to the appropriate SERVQUAL dimensions based on semantic similarity and presence of seed terms. The average polarity was calculated to evaluate the gap between perceived performance and user expectations ([Deshpande, 2025](#)) as shown in Table 2 ([Han et al., 2024](#)), ([Chee Hoo et al., 2024](#)), ([Ye et al., 2024](#)).

Table 2 SERVQUAL Aspects

SERVQUAL Dimension	Aspects
Tangibles	<i>Tampilan aplikasi menarik; tampilan aplikasi tersusun dengan baik; informasi mudah dibaca dan dipahami; navigasi aplikasi jelas; fitur aplikasi mendukung kebutuhan pengguna.</i>
Reliability	<i>Aplikasi berjalan dengan lancar; fungsi aplikasi berjalan sebagaimana mestinya; informasi yang diberikan akurat; aplikasi memberikan layanan secara konsisten; data dan informasi pengguna terlindungi.</i>
Responsiveness	<i>Aplikasi memberikan respons dengan cepat; memberikan informasi terkait status layanan; menyediakan jalur feedback yang jelas; menangani permasalahan pengguna secara aktif.</i>

Assurance	<i>Aplikasi dapat dipercaya; memberikan rasa aman kepada pengguna; informasi yang diberikan dapat dipercaya; menjaga privasi dan keamanan data pengguna; memiliki prosedur layanan yang jelas.</i>
Emphaty	<i>Aplikasi memahami kebutuhan pengguna; memberikan layanan yang sesuai dengan kebutuhan pengguna; menyediakan layanan yang dipersonalisasi; memberikan perhatian terhadap permasalahan pengguna.</i>

The SERVQUAL dimension aspects were manually converted into short seed word phrases, followed by topic mapping (seed mapping), as shown in Table 3.

Table 3 Seed Words GLDA

SERVQUAL Dimensions	Aspects
<i>Tangibles</i>	“tampilan”, “menu”, “fitur”, “ikon”, “desain”, “layout”, “warna”, “font”, “ukuran”, “antarmuka”, “panduan”, “tutorial”, “petunjuk”, “navigasi”, “jelas”, “tidak membingungkan”, “modern”, “estetik”, “tampilan menarik”, “desain bagus”, “layout rapi”, “panduan jelas”, “menu lengkap”, “navigasi mudah”, “font enak dibaca”, “ukuran pas”, “ikon jelas”, “interface bagus”, “aplikasi”
<i>Reliability</i>	“mudah”, “digunakan”, “dipahami”, “tidak error”, “jarang error”, “berfungsi”, “berjalan”, “stabil”, “lancar”, “akses”, “kapan saja”, “dimana saja”, “update”, “terbaru”, “diperbarui”, “akurasi”, “informasi akurat”, “data akurat”, “fungsi normal”, “semua fitur jalan”, “tidak gagal”, “jarang bug”, “informasi valid”, “akses lancar”, “bisa diakses”, “tidak lemot”, “tidak ngefreeze”, “informasi terpercaya”, “update rutin”, “nomor”, “hp”, “mengisi”, “daftar”, “login”, “tidak bisa buka”, “install”
<i>Responsiveness</i>	“cepat”, “respon”, “responsif”, “tanggap”, “dibalas”, “langsung”, “notifikasi”, “keluhan”, “ditanggapi”, “bantuan”, “proses cepat”, “dilayani cepat”, “tindakan cepat”, “respon cepat”, “direspons”, “ditangani segera”, “feedback cepat”, “informasi cepat”, “langsung ditangani”, “langsung dibalas”, “ada pemberitahuan”, “notifikasi update”, “ditangani cepat”, “tanggapan baik”, “langsung diperbaiki”, “fast response”, “notifikasi”
<i>Assurance</i>	“aman”, “keamanan”, “privasi”, “data aman”, “terlindungi”, “otp”, “kode”, “verifikasi”, “kode otp”, “auto debet”, “tagihan akurat”, “tidak bocor”, “tidak salah tagih”, “terpercaya”, “validasi”, “tidak disalahgunakan”, “rahasia”, “reputasi baik”, “data tidak bocor”, “proses aman”, “tidak ada kebocoran”, “keamanan terjaga”, “data pribadi aman”, “login aman”, “akses aman”, “perlindungan data”, “verifikasi aman”, “kode verifikasi”, “kode otp”, “root”, “emulator”.
<i>Empathy</i>	“ramah”, “layanan”, “cs”, “customer service”, “dilayani”, “dibantu”, “memahami”, “mengerti”, “maaf”, “didengarkan”, “diperhatikan”, “nyaman”, “saran”, “dipandu”, “jelas”, “membantu”, “informasi jelas”, “dijelaskan”, “respon ramah”, “tidak dimarahi”, “dihargai”, “diarahkan”, “sabar”, “informasi lengkap”, “penuh perhatian”, “komunikatif”, “mendukung”, “humanis”, “peduli”, “bantu pengguna”, “empati”, “melayani dengan baik”, “ketidaknyaman”, “pintar”, “keluhan”, “pengguna”, “mohon maaf”, “komplain”

2.4.4 Gap Analysis

The formula for SERVQUAL-based gap analysis ([Guo et al., 2024](#)) for service quality evaluation is frequently conducted using the SERVQUAL framework to measure gaps between perceived service performance and user expectations

$$SERQUAL = Performance - Expectation$$

In this study, performance means the proportion of positive sentiment obtained from the SVM results and Expectation means the topic distribution obtained from the GLDA results.

In the SERVQUAL approach, service quality is assessed by comparing customers' expectations with their perceptions of the service received. The analysis can be performed for each SERVQUAL dimension, namely tangibles, reliability, responsiveness, assurance, and empathy ([Mitropoulou & Tsoulfas, 2021](#)):

The calculation formula is:

$$\left(\frac{P1+P2+P3+P4+P5}{5} \right) - \left(\frac{H1+H2+H3+H4+H5}{5} \right) \quad (2)$$

Where P = Perception (performance), H = Expectation, and 5 = Number of sentiments in SERVQUAL dimension

P1 to P5 and H1 to H5 represent the five perception and expectation statements within a particular SERVQUAL dimension. The same calculation is repeated for each remaining dimension.

- Calculation of overall service quality gap (gap measurement)

The calculation formula is:

$$\left(\frac{P1+P2+P3+\dots}{n} \right) - \left(\frac{H1+H2+H3\dots}{n} \right) \quad (3)$$

Where P = Perception, H = Expectation, and n = Total number of reviews

P1 to Pn and H1 to Hn represent all perception and expectation statements, totaling n items. The performance aspects are mapped using a 1–3 Likert scale as shown in Table 4.

Table 4 Perception Scale	
Scale	Label
1	Not Satisfied
2	Moderately Satisfied
3	Satisfied

The classification of expectations based on the Likert scale can be seen in Table 5.

Table 5 Expectation Scale	
Scale	Label
1	Low Expectations
2	Moderate Expectations
3	High Expectations

The results of the gap analysis calculation will be visualized using a spider chart with the help of Microsoft Excel tools.

2.5. Evaluation

The activities conducted in this stage include evaluating sentiment analysis by calculating accuracy, assessing the coherence scores (c_v and UMass) of the topic modeling, and computing the

gap values. The next step calculates the service quality gap based on the SVM and GLDA results.

2.6. Deployment

A simple interactive dashboard was developed using Streamlit to visualize and demonstrate the results of the proposed approach. The dashboard provides a user-friendly interface that allows users to interact with the analysis results without directly accessing or executing the underlying source code. It presents the processed data, model outputs, and relevant evaluation results in an intuitive visual format. A simple dashboard was developed using Streamlit to visualize the results.

3. Results and Discussion

3.1. Business Understanding

Users shared a wide range of experiences using the Mobile JKN app. These insights served as the basis for identifying key service gaps and features needing improvement.

3.2. Data Understanding

User review data was collected via scraping using the `google_play_scraper` library in Python. A total of 44,752 reviews were initially collected, and after removing duplicates, 30,926 unique reviews were retained for further analysis. The scraping results can be seen in Table 6 below.

Table 6 Scraping Result

userName	Content
Pengguna Google	aplikasi tolol server busuk pemerintah korup gini nih klo garap aplikasi
Pengguna Google	simple mudah
Pengguna Google	Bagus
Pengguna Google	Sangat membantu

3.3. Data Preparation

The following preprocessing steps were conducted:

3.3.1 Cleaning

Removed numbers, symbols, whitespace, punctuation, and emoticons. The cleaning results can be seen in Table 7 below.

Table 7 Cleaning Result

Content	Cleaning
Ok banget	Ok banget
Perlu ditingkatkan.....	Perlu ditingkatkan
Akhir2 ini Aplikasi suka tidak mau terbuka	Akhir ini Aplikasi suka tidak mau terbuka
Udah pakai wifi, data seluler selalu ada notif offline	Udah pakai wifi data seluler selalu ada notif offline

3.3.2 Case Folding

Converted all text to lowercase. The case folding results can be seen in Table 8 below.

Table 8 Case Folding Result

Cleaning	Case Folding
Ok banget	ok banget
Perlu ditingkatkan	perlu ditingkatkan
Akhir ini Aplikasi suka tidak mau terbuka	akhir ini Aaplikasi suka tidak mau terbuka
Udah pakai wifi data seluler selalu ada notif offline	udah pakai wifi data seluler selalu ada notif offline

3.3.3 Normalization

Standardized non-formal words using InSet Lexicon dictionary. The normalization results can be seen in Table 9 below.

Table 9 Normalization Result

Case Folding	Normalization
sangat puas dan cepat dgn adanya aplikasi digital bpjs kesehatan semuanya menjadi mudah terimakasih bpjs kesehatan nomer verifikasi ndk muncul udh bulan ndk bisa nomer verifikasinya masuk ke nomer hp padahal msh aktifterlalu bnyk ini itu dipersulit sekali apk bikin emosi tidak bisa daftar dgn alasan nomor sudah terpakai padahal udah nyoba nomor kartu	sangat puas dan cepat dengan adanya aplikasi digital bpjs kesehatan semuanya menjadi mudah terimakasih bpjs kesehatan nomor verifikasi tidak muncul sudah bulan tidak bisa nomor verifikasinya masuk ke nomor hp padahal masih aktif aplikasi membuat emosi tidak bisa daftar dengan alasan nomor sudah terpakai padahal sudah mencoba nomor kartu

3.3.4 Tokenization

Split sentences into individual words. The tokenization result can be seen in Table 10 below.

Table 10 Tokenization Result

Normalization	Tokenization
sangat puas dan cepat dengan adanya aplikasi digital bpjs kesehatan semuanya menjadi mudah terimakasih bpjs kesehatan	['sangat', 'puas', 'dan', 'cepat', 'dengan', 'adanya', 'aplikasi', 'digital', 'bpjs', 'kesehatan', 'semuanya', 'menjadi', 'mudah', 'terimakasih', 'bpjs', 'kesehatan']
aplikasi bodoh server busuk pemerintah korup begini nih kalau garap aplikasi	['aplikasi', 'bodoh', 'server', 'busuk', 'pemerintah', 'korup', 'begini', 'nih', 'kalau', 'garap', 'aplikasi']
bagus sangat membantu	['bagus', 'sangat', 'membantu']

3.3.5 Stopword Removal

Removed common, non-informative words using NLTK. The stopwords removal result can be seen in Table 11 below.

Table 11 Stopword Removal Result

Tokenization	Stopword Removal
['sangat', 'memudahkan', 'kita'] ['lebih', 'mudah', 'dalam', 'menggunakannya']	['sangat', 'memudahkan'] ['lebih', 'mudah', 'menggunakannya']
['sangat', 'bermanfaat', 'bagi', 'kami']	['sangat', 'bermanfaat']

3.3.6 Labeling

Sentiment was labeled using the InSet Lexicon. Labeling was conducted using a 1–3 Likert scale, as shown in Table 1. The labeling result can be seen in Table 12 below.

Table 12 Labeling Result

Tokenization	Label
['sangat', 'puas', 'cepat', 'aplikasi', 'digital', 'bpjs', 'kesehatan', 'mudah', 'terimakasih', 'bpjs', 'kesehatan']	3
['sudah', 'menunggu', 'berjam', 'jam', 'daftar', 'tidak', 'bisa', 'bisa', 'nomor', 'telepon', 'nomor', 'sudah', 'jelek', 'sangat']	1
['down', 'load', 'bisa', 'komen']	2

3.3.7 TF-IDF

Feature extraction was done using Term Frequency-Inverse Document Frequency. An example

of TF-IDF results for the 25th sample can be seen in Table 13 below.

Table 13 TF-IDF Result for the 25th Sample

array position	TF	IDF	TF-IDF	Term
7848	0.08805	10.646238	0.937400	kodeotp
9160	0.08805	3.467692	0.305330	masuk
16124	0.08805	1.902307	0.167498	tidak

3.4. Modeling

3.4.1 Sentiment Analysis using SVM

SVM was used with both linear and RBF kernels, and data was split into 80:20, 70:30, and 50:50 train-test ratios. The numbers of data in the split comparison can be seen in Table 14 below.

Table 14 Total Data in Split Comparison

Split Data	Data Train	Data Test
80:20	24.740	6.186
70:30	21.648	9.278
50:50	15.463	15.463

The accuracy comparison for each data split and kernel type can be seen in Table 15 below.

Table 15 Accuracy Comparison by Split and Kernel

Split Data	Data Train	Data Test	Linear	RBF
80:20	24740	6186	0,95	0,93
70:30	21648	9278	0,94	0,92
50:50	15463	15463	0,94	0,91

Based on Table 15, the highest accuracy was achieved using the linear kernel with an 80:20 split. Therefore, for the next process, the SVM results with the linear kernel and 80:20 split will be used.

3.4.2 Topic Modeling using Guided LDA

Guided LDA was used with seed words derived from SERVQUAL dimensions. Guided LDA was applied using seed words derived from SERVQUAL dimensions. The process involved identifying key aspects from literature (Table 2), converting them into seed words (Table 3), mapping words to topics, extracting features using CountVectorizer, and performing topic modeling with support from the NLTK and Scikit-learn libraries. Topic modelling using GLDA result can be seen in Table 16 below.

Table 16 The Topic Modelling using GLDA Result

Topic	Jumlah	Isi Topik
0	1129	'aplikasi', 'bodoh', 'server', 'busuk', 'pemerintah', 'korup', 'garap', 'aplikasi'
1	894	aplikasi', 'login', 'captcha', 'salah', 'teruspadahal', 'sudah', 'ulangngapain', 'terbitin', 'aplikasi', 'menyusahkan', 'gunain'
2	850	pelayanan', 'tingkatan', 'mohon', 'perbaharui', 'pelayanan
3	1289	'sudah', 'menunggu', 'berjam', 'jam', 'daftar', 'tidak', 'bisa', 'bisa', 'nomor', 'telepon', 'nomor', 'sudah', 'jelek', 'sangat'
4	2024	'sangat', 'puas', 'cepat', 'aplikasi', 'digital', 'bpjs', 'kesehatan', 'mudah', 'terimakasih', 'bpjs', 'kesehatan'

3.4.1 Gap Analysis

Gap analysis aimed to measure the difference between user perception (sentiment-based performance) and expectation using a modified SERVQUAL approach. This involved mapping sentiment classification results to SERVQUAL dimensions. The process included using GLDA topics as SERVQUAL data, encoding and decoding, filtering, and assigning dimension labels. The results are shown in **Table 17**.

Table 17 The results of the SERVQUAL Encoding, Decoding, and Filtering Process

Stopword Removal	SERVQUAL Dimensions
['menyusahkan', 'orang', 'berobat', 'akun', 'persusah', 'verifikasi', 'nomor', 'hp', 'sudah', 'dimasuki', 'kagak', 'masuk']	Reliability
['kode', 'otp', 'tidak', 'masuk', 'masuk', 'sudah', 'kirim', 'ulang', 'kali', 'tidak', 'otp', 'masuk']	Assurance
['mobile', 'jkn', 'tidak', 'bisa', 'logintolong', 'solusi']	Tangible
['sangat', 'bermanfaat']	Responsiveness
['pelayanan', 'tingkatan', 'mohon', 'perbaharui', 'pelayanan']	Empathy
['pokoknya', 'mantaplah']	No Match

The distribution of review counts across SERVQUAL dimensions can be seen in Table 18 below.

Table 18 Distribusi Dimensi SERVQUAL

SERVQUAL Dimensions	Total
Tangibles	1252
Reliability	886
Responsiveness	702
Assurance	986
Empathy	1506
No Match	854

Based on the satisfaction categories in Table 1, the performance score was calculated by multiplying the number of dimensions by the corresponding Likert scale. The distribution of dimensions across performance scales is presented in Table 19 below

Table 19 Distribution of Dimensions Across Performance Scales

SERVQUAL Dimensions	Not Satisfied (1)	Moderately Satisfied (2)	Satisfied (3)	Total
Tangibles	809	287	156	1252
Assurance	819	124	43	986
Empathy	76	336	1094	1506
Reliability	627	185	74	886
Responsiveness	278	193	231	702
No Match	186	475	193	854

Performance aspect calculations were carried out by multiplying the number of data points in each SERVQUAL dimension by the performance scale described in Table 19. The results of the performance score calculation are presented in Table 20 below.

Table 20 Performance Aspect Calculation

SERVQUAL Dimensions	Not Satisfied (1)	Moderately Satisfied (2)	Satisfied (3)	Total	Dimensions Score
Tangibles	809	574	468	1851	
Assurance	819	248	129	1196	
Empathy	76	672	3282	4030	9653
Reliability	627	370	222	1219	
Responsiveness	278	386	693	1357	
Total	809	574	468	1851	19306
Nilai Kinerja		9653			

Table 20 shows performance score calculations; for example, in the Tangibles dimension, scores are derived by multiplying the number of responses by their Likert scale (e.g., $809 \times 1 = 809$). Table 21 presents expectation scores, calculated by multiplying the total data per dimension by the highest Likert value (3), assuming users generally have high expectations.

Table 21 The Result of The Expectation Score

SERVQUAL Dimensions	Total	Expectation Score	Score
<i>Tangibles</i>	1252	3	3756
<i>Assurance</i>	986	3	2958
<i>Empathy</i>	1506	3	4518
<i>Reliability</i>	886	3	2658
<i>Responsiveness</i>	702	3	2106
Total		15996	

Two types of gap analysis were conducted: dimension-by-dimension and overall SERVQUAL.

- *Dimension by Dimension*

After calculating the performance and expectation scores, the dimension-by-dimension gap was computed using the following formula:

- *Tangible*

$$Gap = \left(\frac{1851}{3756} \right) - \left(\frac{3756}{3756} \right) = 0,49 - 1 = -0,51 = -51\%$$

○ Assurance

$$Gap = \left(\frac{1196}{2958}\right) - \left(\frac{2958}{2958}\right) = 0,40 - 1 = -0,6 = -6\%$$

○ Empathy

$$Gap = \left(\frac{4030}{4518}\right) - \left(\frac{4518}{4518}\right) = 0,89 - 1 = -0,11 = -11\%$$

○ Reliability

$$Gap = \left(\frac{1219}{4518}\right) - \left(\frac{4518}{4518}\right) = 0,26 - 1 = -0,74 = -74\%$$

○ Responsiveness

$$Gap = \left(\frac{1357}{2106}\right) - \left(\frac{2106}{2106}\right) = 0,64 - 1 = -0,36 = -36\%$$

Based on the gap calculation, the Tangible dimension showed a gap of -51%, indicating a 51% shortfall in meeting user satisfaction. The Assurance dimension had a gap of -6%, the Empathy dimension -11%, the Reliability dimension -74%, and the Responsiveness dimension -36%. Each reflecting the percentage needed to reach user satisfaction in their respective areas.

The reliability dimension shows the largest gap (-0.74), indicating that users frequently experience system-related problems such as login failures, application errors, and server instability. This finding is consistent with the sentiment analysis results, where many reviews contain terms such as “error”, “login”, and “cannot access”. These results suggest that technical stability remains one of the main issues affecting user satisfaction with the Mobile JKN application.

The results indicate that several service quality dimensions show significant gaps between user expectations and perceived performance. In particular, the reliability dimension shows the largest negative gap, indicating that users frequently experience issues related to application stability and system performance. These findings suggest that improving system reliability and responsiveness should become a priority for developers in order to enhance user satisfaction with the Mobile JKN application.

• SERVQUAL

Following the dimension-by-dimension analysis, the overall SERVQUAL score was calculated using the following formula:

$$SERVQUAL \text{ Based Gap Analysis} = \left(\frac{19306}{30926}\right) - \left(\frac{15996}{30926}\right) = 0,62 - 0,51 = 0,11 = 11\%$$

Based on the results, the overall SERVQUAL score was 11%, indicating that user satisfaction with the Mobile JKN application is only 11%.

3.5. Evaluation

3.5.1 SVM

SVM accuracy across various split scenarios and kernel types is shown in Table 15. The highest accuracy was achieved using the linear kernel with an 80:20 split, reaching 0.95. The confusion matrix evaluation for the linear kernel with the 80:20 split is presented in Figure 1.

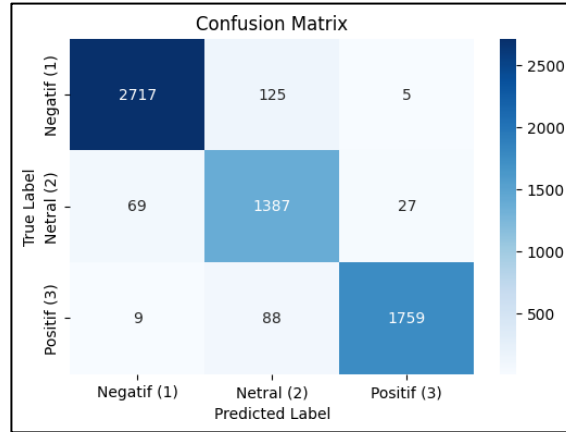


Figure 2 Confussion Matrix 80:20 Linear Kernel

Figure 2 shows the confusion matrix results: 2,717 true negatives, 1,387 true neutrals, and 1,759 true positives. Misclassifications include 130 false negatives, 96 false neutrals, and 97 false positives. Confusion Matrix for RBF kernel by split 80:20 can be seen in Figure 2.

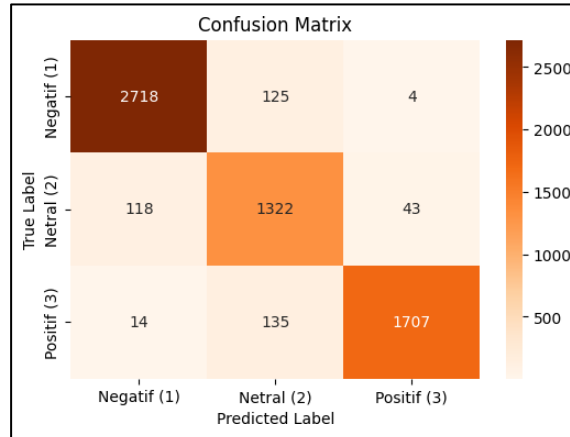


Figure 3 Confusion Matrix 80:20 RBF Kernel

Figure 3 presents the confusion matrix for the RBF kernel (80:20 split), showing 2,718 true negatives, 1,322 true neutrals, and 1,707 true positives. Misclassifications include 129 false negatives, 161 false neutrals, and 149 false positives. The classification report for the linear kernel with an 80:20 split is presented in Table 21 below.

Table 21 Classification Report 80:20 Linear Kernel				
Class	Precision	Recall	F-1 Score	Support
Negative	0.97	0.95	0.96	2847
Neutral	0.87	0.94	0.90	1483
Positive	0.98	0.95	0.96	1856
Macro Avg	0.94	0.95	0.92	6186
Accuracy			0.95	6186

Negative refers to Likert scale 1 (Not Satisfied), Neutral refers to Likert scale 2 (Moderately Satisfied), and Positive refers to Likert scale 3 (Satisfied).

Table 21 shows that the SVM model performs consistently across all three classes, with high scores particularly in the positive and negative classes: precision scores of 0.98 for positive and 0.97 for negative, recall scores of 0.95 for both, and F1-scores of 0.96 for both classes. The neutral class

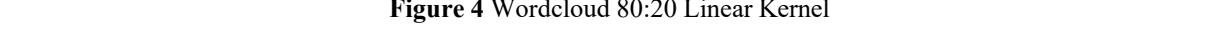
Table 22. Chloride ion in DMSO at 298.15 K. 1BDE

Class	Precision	Recall	F-1 Score	Support
-------	-----------	--------	-----------	---------

These findings are consistent with previous studies that reported strong performance of SVM

The experimental results show that the SVM model using the linear kernel achieved bet

WordCloud is a visualization of the most frequently occurring words in user reviews of the



- "*tidak*" (not) and "*bisa*" (can't): indicate user complaints related to login failures, errors, or inability to perform actions.

- "*fasilitas*" (facilities), "*kesehatan*" (health), and "*rumah sakit*" (hospital): point to user feedback on healthcare services, the core function of the app.
- "error", "*masuk*" (login), and "*daftar*" (register): suggest technical issues like login and registration failures.

3.5.2 GLDA Evaluations

The Guided Latent Dirichlet Allocation (GLDA) model was evaluated using coherence c_v and UMass scores. The c_v coherence score for the five analyzed topics was 0.64, indicating a reasonably good level of word association within the topics. Although not close to 1, this value suggests that the topics are sufficiently coherent and interpretable.

The UMass score was -1.99. As UMass values are typically negative, scores closer to 0 (from the negative side) reflect better coherence. Therefore, a score of -1.99 indicates that five is an optimal number of topics for topic modeling using GLDA on Mobile JKN app reviews.

3.5.3 Gap Analysis Evaluations

The dimension-by-dimension gap is visualized using a spider chart, as shown in Figure 4 below.

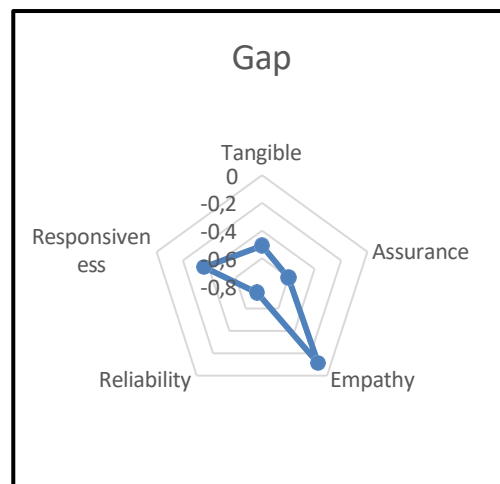


Figure 5 Dimensions by Dimensions Visualisation

Figure 5 presents the dimension-by-dimension gap analysis. The Assurance dimension shows the smallest gap at -6%, indicating minimal discrepancy between performance and user expectations. In contrast, the Reliability dimension has the largest gap at -74%, highlighting a critical area that requires improvement.

The overall SERVQUAL gap value is 0.11 (11%), indicating a moderate discrepancy between user expectations and perceived service performance. This gap suggests that improvements are still needed to fully align the Mobile JKN application's performance with user expectations.

3.6. Deployment

A simple dashboard was developed using Streamlit to visualize the results as shown in Figure 4.

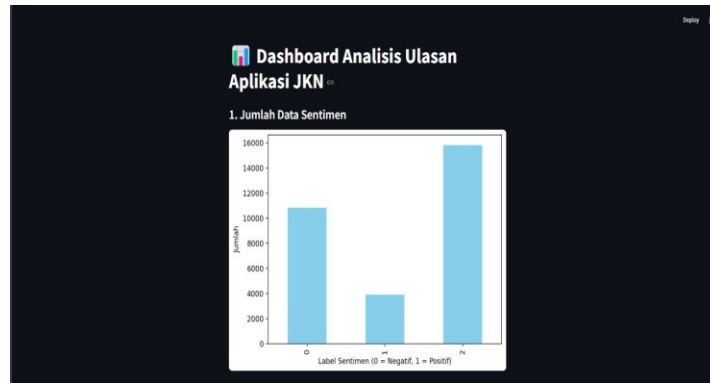


Figure 6 Dashboard Streamlit

The deployment process was carried out using the Streamlit framework (Figure 4), with all programming implemented in Visual Studio Code (VS Code). The resulting dashboard presents several key analytical outcomes. First, it displays the sentiment distribution generated from the Support Vector Machine (SVM) model, showing the classification of user reviews into negative, neutral, and positive sentiments. Second, it visualizes the topic distribution produced by the Guided Latent Dirichlet Allocation (GLDA) model across the Mobile JKN user reviews. Lastly, it presents the SERVQUAL dimension mapping, showing how the reviews are categorized into dimensions such as tangibles, reliability, responsiveness, assurance, and empathy.

4. Conclusions

This study concludes the following key findings. Sentiment analysis using Support Vector Machine (SVM) with a linear kernel and an 80:20 data split revealed that most user reviews of the Mobile JKN application were negative (2,784 reviews), followed by positive (1,791) and neutral (1,611) sentiments. Similarly, the RBF kernel with the same split showed a dominant number of negative sentiments (2,847), followed by positive (1,856) and neutral (1,483) reviews. These results suggest a general tendency toward dissatisfaction among users, indicating the need for service improvement.

Topic modeling with Guided Latent Dirichlet Allocation (GLDA) successfully categorized the reviews into five main SERVQUAL-based dimensions: tangible, assurance, empathy, reliability, and responsiveness. The empathy dimension was the most frequently mentioned, while responsiveness appeared the least. Coherence scores of 0.64 (c_v) and -1.99 (UMass) demonstrate that the topics generated are coherent and interpretable.

The gap analysis showed that assurance had the smallest gap at -6%, while reliability had the largest at -74%. The overall SERVQUAL gap score was 11%, suggesting that while general service quality was moderately aligned with user expectations, significant improvements are still needed in specific areas, particularly reliability.

5. Limitations

This study has several limitations. First, sentiment labeling relies on lexicon-based preprocessing, which may introduce classification bias. Second, the analysis only uses reviews from the Google Play Store and may not represent all Mobile JKN users. Third, although SVM provides strong performance for text classification, the model does not capture contextual semantics as effectively as deep learning approaches such as transformer-based models.

6. Contribution

This research contributes by proposing an integrated analytical approach that combines sentiment

classification, topic modeling, and SERVQUAL-based gap analysis to evaluate mobile service quality using user-generated reviews. The study demonstrates how machine learning techniques can support data-driven service improvement in digital public health applications.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. I Kadek Dwi Nuryana, S.T., M.Kom., the academic supervisor, for his guidance and valuable support throughout the completion of this research.

Conflict of Interest

The authors declare no conflicts of interest.

Author Contributions

Conceptualization [BNAS and IKDN]; Methodology, [BNAS]; Software, [BNAS]; Data curation, [BNAS]; Formal analysis, [BNAS]; Investigation, [BNAS]; Visualization, [BNAS]; Supervision, [IKDN]; Writing – original draft, [BNAS]; Writing – review & editing, [IKDN and MNPM] All authors have read and agreed to the published version of the manuscript.

Ethics Statement

This study used Mobile JKN reviews data freely accessed on Google Play Store. The study did not involve direct recruitment or interaction with human participants. Ethical approval was not applicable to the data analysis conducted in this study.

AI Use Declaration

The authors used ChatGPT (OpenAI) as a generative AI-assisted tool during the preparation of this manuscript, primarily for language editing, grammatical refinement, and improving the clarity and organization of the manuscript. The authors reviewed and verified all AI-assisted content and take full responsibility for the accuracy, originality, interpretation, and final content of the published work.

References

- Alanzi, T. M. (2022). Users' satisfaction levels about mHealth applications in post-Covid-19 times in Saudi Arabia. *PLoS ONE*, 17(5 May). <https://doi.org/10.1371/journal.pone.0267002>
- Alemerien, K., Al-Ghareeb, A., & Alksasbeh, M. Z. (2024). Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization. *Journal of Mobile Multimedia*, 20(5), 1089–1116. <https://doi.org/10.13052/jmm1550-4646.2055>
- Barik, K., & Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-023-00861-x>
- Bichri, H., Chergui, A., & Hain, M. (2024). Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 15, Number 2). www.ijacsa.thesai.org
- Chee Hoo, W., Madhavedi, S., Kinkar Pandey, R., Ee Tzo, L., Kai Wah, C., Yan, X., & Prompanyo, M. (2024). Customer satisfaction through service quality dimensions: A study on perishable products purchased through online mobile apps in Klang Valley, Malaysia. *Journal of Infrastructure Policy and Development*, 8(15), 8791. <https://doi.org/10.24294/jipd8791>
- Deshpande, S. K. (2025). SERVICE QUALITY, CUSTOMER SATISFACTION, AND LOYALTY IN EVENT PLANNING: A STUDY OF INDIAN WEDDINGS. *International Journal of Research -GRANTHAALAYAH*, 13(4). <https://doi.org/10.29121/granthaalayah.v13.i4.2025.6130>
- Giebel, G. D., Speckemeier, C., Schrader, N. F., Abels, C., Plescher, F., Hillerich, V., Wiedemann, D.,

- Börchers, K., Wasem, J., Blase, N., & Neusser, S. (2024). Quality assessment of mHealth apps: a scoping review. In *Frontiers in Health Services* (Vol. 4). Frontiers Media SA. <https://doi.org/10.3389/frhs.2024.1372871>
- Guo, T., Li, T., & Qi, Z. (2024). Perceived school service quality and vocational students' learning satisfaction: Mediating role of conceptions of vocational education. *PLoS ONE*, 19(8). <https://doi.org/10.1371/journal.pone.0307392>
- Han, T., Wei, Q., Wang, R., Cai, Y., Zhu, H., Chen, J., Zhang, Z., & Li, S. (2024). Service Quality and Patient Satisfaction of Internet Hospitals in China: Cross-Sectional Evaluation With the Service Quality Questionnaire. *Journal of Medical Internet Research*, 26. <https://doi.org/10.2196/55140>
- Haoues, M., Mokni, R., & Sellami, A. (2023). Machine learning for mHealth apps quality evaluation: An approach based on user feedback analysis. *Software Quality Journal*, 31(4), 1179–1209. <https://doi.org/10.1007/s11219-023-09630-8>
- Kermanshachi, S., Nipa, T. J., & Nadiri, H. (2022). Service quality assessment and enhancement using Kano model. *PLoS ONE*, 17(2 February). <https://doi.org/10.1371/journal.pone.0264423>
- Li, C., Li, S., Yang, J., Wang, J., & Lv, Y. (2023). Topic evolution and sentiment comparison of user reviews on an online medical platform in response to COVID-19: taking review data of Haodf.com as an example. *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1088119>
- Mitropoulou, A. D., & Tsoulfas, G. T. (2021). Using a Modified SERVQUAL Approach to Assess the Quality of Supply Chain Services in Greek Online Supermarkets. *Logistics*, 5(4). <https://doi.org/10.3390/logistics5040069>
- Ossai, C. I., & Wickramasinghe, N. (2023). Sentiments prediction and thematic analysis for diabetes mobile apps using Embedded Deep Neural Networks and Latent Dirichlet Allocation. *Artificial Intelligence in Medicine*, 138. <https://doi.org/10.1016/j.artmed.2023.102509>
- Qiu, Y., Xiao, C., & Li, J. (2024). Service quality evaluation of county-level public hospitals in Chongqing under smart healthcare. *BMC Health Services Research*, 24(1). <https://doi.org/10.1186/s12913-024-11976-z>
- Rahardi, M., Aminuddin, A., Abdulloh, F. F., & Nugroho, R. A. (n.d.). Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia. In *IJACSA International Journal of Advanced Computer Science and Applications* (Vol. 13, Number 6). Retrieved <https://t.co/h5x41UO3tF>
- Sanwal, M., & Muhammad, M. M. (2023). Performance Comparison of Machine Learning and Deep Learning Models for Sentiment Analysis of Hotel Reviews. *International Journal of Information Technology and Applied Sciences (IJITAS)*, 5, 1-7.
- Sharifi, T., Hosseini, S. E., Mohammadpour, S., Javan-Noughabi, J., Ebrahimipour, H., & Hooshmand, E. (2021). Quality assessment of services provided by health centers in Mashhad, Iran: SERVQUAL versus HEALTHQUAL scales. *BMC Health Services Research*, 21(1). <https://doi.org/10.1186/s12913-021-06405-4>
- Venugopalan, M., & Gupta, D. (2022). An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis. *Knowledge-Based Systems*, 246. <https://doi.org/10.1016/j.knosys.2022.108668>
- Ye, L., Yang, S., Jiang, B., Liu, C., Jin, X., & Chang, P. (2024). Developing a quality indicator system for evaluating internet plus home care nursing services based on the SERVQUAL model: a Delphi-analytic hierarchy process study. *PeerJ*, 12(11). <https://doi.org/10.7717/peerj.18281>
- Zaim Mohd Salleh, M., Sara Nadia Muhamad Yunus, N., Maisarah Abdul Ghani, N., Saiful Bahry, N., Mohd Ali, A., & Author, C. (2024). Exploring E-Customer Satisfaction Using e-SERVQUAL Among Online Shoppers. In *Information Management and Business Review* (Vol. 16, Number 3S).