IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 12, No. 4, December 2023, pp. 1947~1959
ISSN: 2252-8938, DOI: 10.11591/ijai.v12.i4.pp1947-1959

Using machine learning to improve a telco self-service mobile application in Indonesia

Jwalita Galuh Garini, Achmad Nizar Hidayanto, Agri Fina
Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia

Article Info

Article history:

Received Nov 24, 2022
Revised Jan 11, 2023
Accepted Mar 10, 2023

Keywords:

Latent dirichlet allocation
Machine learning
Self-service mobile application
Sentiment analysis
Service quality
Topic modeling

ABSTRACT

The use of mobile applications extends to the telecommunication sector, mainly due to COVID-19. Failure to provide it can cause dissatisfaction and result in the removal of the mobile application. Moreover, this leads to lost service opportunities, so paying attention to the mobile application's quality is essential. There has yet to be a study on measuring the service quality of a self-service mobile application in the telecommunication sector using online customer reviews. This study uses sentiment analysis and topic modeling to determine the service quality of a self-service mobile application in the telecommunication sector from reviews on Google Play Store and Apple App Store. This study uses myIndiHome as a case study. The total data obtained from both platforms are 20,452 reviews. Sentiment analysis was performed using Naïve Bayes, support vector machine, and logistic regression, while topic modeling was performed using latent dirichlet allocation. The results show that logistic regression performs better than support vector machine and Naïve Bayes. Meanwhile, topic modeling shows that the positive review data has three topics, including application features, products/services, and application interfaces. Moreover, the negative review data has five topics, including application availability, application feature reliability, application processing speed, bugs, and application reliability.

Corresponding Author:

Jwalita Galuh Garini
Faculty of Computer Science, University of Indonesia
Jakarta, Indonesia
Email: jwalita.galuh@ui.ac.id

1. INTRODUCTION

Technological developments in the digital era are driving lifestyle changes to be more practical. Mobile applications are now an integral part of everyday life. Mobile application users spend an average of 4-5 hours daily using mobile apps [1]. In 2021, the global mobile application market generated 230 billion downloads, while Indonesia is the country with the fifth largest mobile application market globally, generating 7.31 billion downloads by 2021 [1]. The growth in the number of downloads in Indonesia is also one of the largest in the world, increasing by 15% compared to 2020 [1]. As smartphone users continue to increase, the number of application downloads is projected to increase [2], [3]. As of the second quarter of 2022, there are more than 3.5 million mobile applications on Google Play Store, 2.2 million mobile applications on Apple App Store, and 515 thousand million mobile applications on other platforms [3], [4].

The use of this mobile application also extends to the telecommunications sector, mainly due to COVID-19, where many users are increasingly developing and using self-service mobile applications. A self-service mobile application for telecommunication service users becomes a communication link platform between users and companies. This mobile application allows users to have complete control over their services anywhere and anytime, from adding services, checking bills, service information and promos, to complaints.

The use of mobile applications will have an impact on the business success, failure to deliver may cause user dissatisfaction and result in removing the mobile application, resulting in lost service opportunities [5]. So in providing a mobile application, it is essential to pay attention to the quality of the mobile application.

In the competitive mobile application industry, user experience has an essential role in the market penetration of mobile applications [6]. Users' opinions about an application can define the perceived quality of the mobile application, which can be demonstrated through feedback in the form of reviews and ratings [7]. Mobile application reviews can provide valuable information for users to discover what others think about mobile applications. It is also valuable for mobile application providers to receive user feedback about features they like or expect and bugs in mobile applications [7].

Research on service quality in mobile applications has been carried out before. Leem and Eum [8] measure service quality and detect complaints on mobile banking applications by sentiment analysis and topic modeling against review data from the Google Play Store. From sentiment analysis, it was obtained that most reviews were positive. While from topic modeling on negative reviews, it was known that the topics of customer complaints related to technology, interaction, customer convenience, and process [8]. Meanwhile, Oyebode *et al.* [9] evaluated mental health mobile applications with sentiment analysis and thematic analysis on reviews from Google Play Store and Apple App Store. This research obtains the factors that positively and negatively influence the effectiveness of mental health mobile applications and become recommendations for developers to increase the effectiveness of mobile applications [9]. Nayebi *et al.* [10] also did the same for 70 top charts of mobile applications. It analyzed the possibility of merging sources of information between reviews from the Google Play Store and content from Twitter to support mobile application development. This research shows that when conducting empirical studies on user feedback for mobile application quality assessment, one should also look to additional sources of information [10].

As mentioned earlier, several studies have used sentiment analysis and topic modeling to measure service quality on mobile applications [8]–[10]. However, there has yet to be any similar study on self-service mobile applications in the telecommunications sector. Before, Bhale and Bedi [5] conducted a qualitative study based on survey data to measure the level of customer engagement and satisfaction with service channels and the reasons for dissatisfaction using digital self-service by telecommunications consumers. Several factors were found to be reasons for dissatisfaction with self-service mobile applications, including application speed, unwanted information, incomplete information, unavailable information/services, application response failure rate, and difficulty navigating. Collecting samples using surveys requires much time and effort, so it is considered less effective. Data survey is starting to be replaced with data from online interactions. Therefore, this study will measure the service quality of self-service mobile applications in the telecommunications sector by utilizing reviews from Google Play Store and Apple App Store, also the text mining approach.

Mobile application reviews can be used as a reference for making business decisions according to the results of an analysis of user opinions about the product or service used [11]. While the combination of sentiment analysis and topic modeling can help understand the sentiment and find topics being discussed on a related platform to be used in developing strategies to improve certain services or products [12]. There are several main approaches that are commonly used for sentimeUbnt analysis and topic modeling. Machine learning is a widely used sentiment analysis approach because of its simple algorithm and high classification accuracy [13]–[15]. Moreover, latent dirichlet allocation is the most frequently used method for modeling topics because it is very suitable for modeling general topics using various data [16], [17]. Because of these characteristics, this study will use these methods to answer research questions. This study also tried to use n-grams as a feature in an experiment to see feature size on classification performance.

This study wants to answer two research questions. The research questions are "how is the sentiment towards self-service mobile applications in the telecommunication sector?" and "what recommendations can be made to improve the service quality of self-service mobile applications based on the topics obtained from negative sentiment?". This study is also organized into five sections. Section 1 describes the introduction and research background, Section 2 describes the study of relevant literature, Section 3 describes the process used in this study, Section 4 describes the results and discussion, and Section 5 describes the conclusions.

2. LITERATURE STUDY

2.1. Mobile application service quality

Service quality is an essential topic in the traditional service industry to the mobile service industry [8]. Service quality is also an essential factor affecting customer satisfaction and loyalty and determining service providers' success [18]. Many studies have examined service quality, and one of the most influential is the research by Parasuraman *et al.* [19], which developed the service quality (SERVQUAL) instrument. This model includes reliability, assurance, tangibles, empathy, and responsiveness [19]. The initial concept of evaluating service quality was inadequate for virtual environments where customers interact with technology rather than people, so the electronic service quality (E-S-QUAL) model was developed. It measures service quality in an electronic or website environment and consists of efficiency, system availability, fulfillment, and privacy [20]. Furthermore, Huang *et al.* developed the mobile service quality (M-S-QUAL) to measure service quality in a mobile environment [21]. The dimensions of the M-S-QUAL model consist of contact, fulfillment, privacy, efficiency, and responsiveness [21].

As smartphones and mobile applications grow, service delivery demands new aspects to be considered [18]. Service quality perceived by mobile application users needs to consider service design requirements [18]. So, Wulfert [18] proposed the mobile application service quality model. The difference between M-S-QUAL and mobile application service quality lies in service provision. M-S-QUAL focuses on services accessed through mobile devices, while mobile application service quality focuses more on mobile applications that run on mobile devices to provide mobile services to users [18]. Simply, the mobile application service quality can be shown as a subset of the M-S-QUAL. The mobile application service quality consists of two types of dimensions, namely primary and secondary, described in Figure 1. Details of each dimension are explained in Table 1.

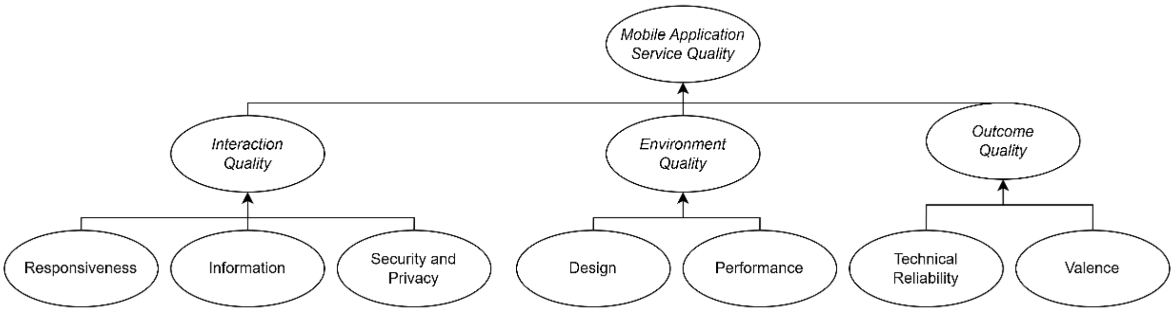

Figure 1. Hierarchy of mobile application service quality

Table 1. Wulfert's mobile application service quality dimensions

Dimension	Description	Item
Interaction quality	Shows all the quality characteristics of interaction between customers and service providers	
Responsiveness	The ability of service providers to promptly and politely resolve customer issues related to mobile applications	Availability of customer service Ability to solve problems Personel's politeness and kindness Availability of mobile application's guidance and instructions for use
Information	The service provider provides accurate and precise information	Adequacy of information Use of information Correctness of information
Security and privacy	System and network resources protection from any attack from external or internal along with customer personal data protection	Information security Data protection Data collection
Environment quality	Shows the context of mobile application delivery and the quality characteristics of equipment that affect mobile application delivery	
Design	Aesthetic features and user interface design layout	Visual aesthetics and layout clarity Multimedia content quality Ease of use and navigation covenience Search and filter function
Performance	Mobile application performance and resource requirements	Speed of processing Usage of storage device and usage of mobile network Quality of device and connection
Outcome quality	Shows the technical quality of service delivery and customer satisfaction towards mobile services	
Technical reliability	Operation of mobile applications and services provided accurately and consistently	Reliability of mobile applications and features Availability of mobile services Continuous operation performed after the update
Valence	Customer's final impression after the completion of service delivery	Overall satisfaction towards mobile services Satisfaction with the scope of services

2.2. Sentiment analysis

Sentiment analysis can be defined as a study that analyzes opinions, sentiments, judgments, attitudes, and emotions from people toward entities in the form of products, services, organizations, individuals, events, issues, or topics [22]. It is expressed in written text as positive, neutral, or negative [22]. This study can be applied to many domains, such as consumer products, healthcare, tourism, hospitality, financial services, social events, and political elections [22]. Researchers, business organizations, and governments alike can use sentiment analysis to analyze public emotions and views to get business insights and create better decisions [15]. Several main approaches for sentiment analysis are commonly used, including machine learning, lexicon based, and hybrid [14], [15]. Supervised machine learning is an approach that is widely used in this research field due to its simple algorithm and high accuracy results, with support vector machine and Naïve Bayes widely used as primary methods [14], [15].

Support vector machine is a non-probabilistic classifier that can split data linearly or non-linearly and handle discrete and continuous variables [15]. Support vector machine is included in the kernel methods category, an algorithm that relies on data only through the dot-product or can be replaced by a kernel function [23]. This classifier analyzes data and finds the optimal hyperplane to separate data into different classes [15]. The separation between hyperplanes is set to be as large as possible [24]. Effective separation is demonstrated by the hyperplane having the maximum margin to the closest training point of the two classes [15].

Naïve Bayes is a simple probabilistic classifier based on Bayes' Theorem and relies on bag of words (BoW) feature extraction [14], [15]. Naïve Bayes predicts the probability of a specific group of features as part of a specific label [14]. The words' position in the document is ignored, and a specific word's existence is independent of the existence of other words [15]. Naïve Bayes assigns document D to category C, which maximizes the value of P(C|D) by applying Bayes' rule [15]. It shows how often category C happens given that document D happens, written as P(C|D), when we know how often document D happens given that category C happens, written as P(D|C), and how likely C and D are on their own, written as P(C) and P(D).

Logistic regression is another machine learning method for a classification task. Logistic regression works by multiplying the input value with the weight value [14]. It estimates the probability of a discrete outcome based on a given input variable. This classifier learns which input property is most helpful for identifying classes [14]. To calculate the best parameter, logistic regression uses maximum-likelihood [14].

2.3. Topic modeling

Topic modeling is an approach to presenting hidden concepts from large volumes of data [16]. Topic modeling is an algorithm for finding and annotating extensive collections of documents [25]. The topic modeling algorithm is a statistical method for analyzing the words in a text to find the topics within it and how they are connected and change over time [25]. This algorithm requires no prior labeling as topics arise from an analysis of the original text, thus enabling us to manage electronic records on a scale that is impossible with manual labeling [25]. This algorithm can be applied to many data types, some have used it to find patterns in genetic data, images, and social networks [25].

Latent dirichlet allocation is the simplest and most commonly used method for finding topics in text documents [25], [26]. This method was developed to fix problems in the probabilistic latent semantic analysis, which is a probabilistic version of latent semantic analysis [25]. The basic idea of latent dirichlet allocation is that documents are represented as a random mix of hidden topics, where each topic is characterized by a word distribution [26]. Latent dirichlet allocation assumes that topics are generated before documents. So for each document, the word is generated through two stages [25]. First, randomly selected distribution of topics. Second, for each word in the document, randomly select topics from the distribution over the topics in step 1, then randomly select a word from the appropriate vocabulary distribution.

3. METHOD

This study aims to determine the service quality of a self-service mobile application in the telecommunications sector using sentiment analysis and topic modeling based on reviews from Google Play Store and Apple App Store. This study uses myIndiHome as a case study because it is one of the self-service mobile applications in the telecommunication sector with the most users in Indonesia. Currently, myIndiHome application for Android users has been reviewed 160 thousand times, and for iOS users has been reviewed 8.2 thousand times [27], [28]. The method used in this study is quantitative with the type of mono-method quantitative study, or quantitative research using a single data collection technique. Figure 2 shows the outline of the stages carried out in this study based on the general text analysis framework [29].

3.1. Data collection

Data was collected from myIndiHome mobile application review column on Google Play Store and Apple App Store. It was scraped using the Python library, namely Google Play Scraper and Apple App Store Scraper. Review data was collected starting November 1, 2021, from the launch of myIndiHome application version 4 nationally, until April 30, 2022. Part of the collected data will be manually annotated by three researchers and used to train and evaluate classifier models.

3.2. Data preprocessing

The data obtained needs to be appropriately prepared to be used as input for the data mining algorithm [30]. This process aims to transform the semi-structured text and unstructured text into a structured vector space model [31]. In other words, data preprocessing converts real-world raw data into a computer-readable format [30]. The initial data processing steps in this study follow the primary stages [31]:

- Case folding: converts all characters in the review data to lowercase.
- Cleansing: removes non-American standard code for information interchange (ASCII) characters, uniform resource locators (URL) addresses, hashtags, punctuation, numbers, new lines, and extra spaces.
- Tokenization: dividing the existing text into smaller and meaningful elements, in this study, splitting sentences into words.
- Normalization: change slang words into standard words using a slang word dictionary in Indonesian.
- Stopping removing stopwords from the text so we can focus more on essential words, in this study, using Sastrawi and natural language toolkit (NLTK) libraries in Python.
- Stemming: changing words into forms without affixes, in this study, using Sastrawi library in Python.

3.3. Sentiment analysis

The model for classifying sentiments was trained using supervised machine learning, which consists of several popular classifiers, including Naïve Bayes, support vector machine, and logistic regression. The classifier model will be evaluated using measurement of accuracy, precision, recall, and F-score, also using K-fold cross-validation with 10-fold. The best classifier model obtained at this stage will be used to identify sentiments in the entire data.

3.4. Topic modeling

The review data obtained from the previous stage will be used for topic modeling based on positive and negative sentiments at this stage. The goal is to find out the main discussion topics for each sentiment. Topics that users like can be obtained from positive sentiment reviews, while topics that users complain about can be obtained from negative sentiment reviews. This study uses latent dirichlet allocation algorithm for topic modeling. The number of topics is determined by calculating the highest coherence value.

3.5. Opportunity for improvement

The results obtained from sentiment analysis and topic modeling will be analyzed further. This analysis aims to find opportunities to improve mobile self-service applications in the telecommunications sector. The results can be used as recommendations for application providers in providing self-service mobile applications according to user needs.

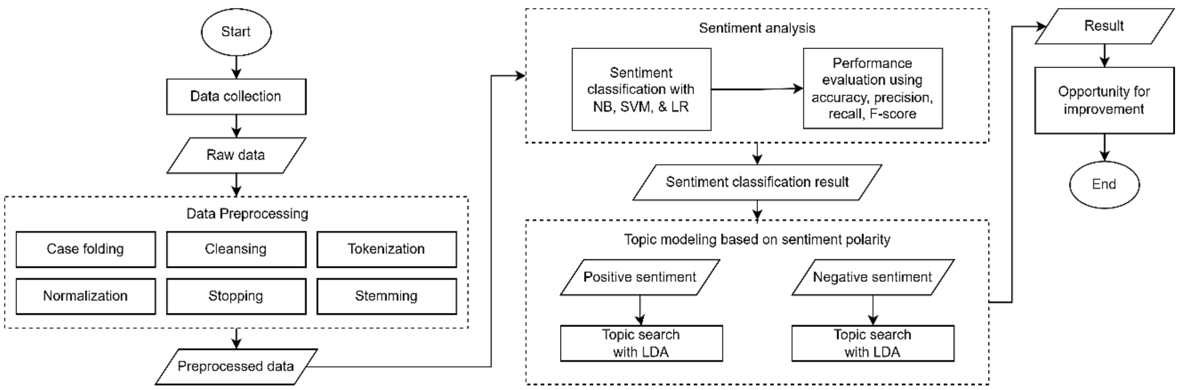

Figure 2. Research process

4. RESULTS AND DISCUSSION

This section consists of three sub-sections: results, discussion, and implication. The results sub-section will explain the results obtained based on the methodology described in Section 3. Meanwhile, the discussion sub-section will explain the interpretation of the results obtained. Moreover, the implication sub-section will explain the theoretical and practical implications of the research conducted.

4.1. Results

4.1.1. Data collection

Data were obtained from Google Play Store and Apple App Store using the Python library of 19,211 and 1,241 reviews related to the mobile self-service application in the telecommunications sector in Indonesia, namely myIndiHome. So that the total data obtained from the two platforms is 20,452 reviews. From this data, a sample of 2,045 reviews was taken to annotate sentiment manually. The manual annotation process generated 871 positive reviews, 300 neutral reviews, and 874 negative reviews.

4.1.2. Data preprocessing

The data preprocessing consists of case folding, cleansing, tokenization, normalization, stopping, and stemming using Python libraries, namely natural language toolkit (NLTK) and Sastrawi. Before data preprocessing, the average number of words in each review is 16 words. Meanwhile, after data preprocessing, the average number of words in each review is 9 words. The results of the data preprocessing left a total of 19,783 reviews.

4.1.3. Sentiment analysis

This study used 2,045 manually annotated reviews to train the model. It also divides sentences into word parts used for modeling, consisting of unigram, bigram, and trigram. Meanwhile, the classifiers used consist of Naïve Bayes, support vector machine, and logistic regression. For model testing, 10-fold cross-validation is used. Data is divided into ten equal parts, with 9-fold used as the train data for each iteration, while the remaining folds are used as test data. The best classifier model is obtained by comparing accuracy, precision, recall, and F-score values based on algorithm variables and n-grams. The measurement values obtained for each model are shown in Table 2.

Table 2. Classifier performance results

Classifier	n-gram	Accuracy	Precision	Recall	F-score
Naïve bayes	Unigram	79.02%	52.68%	61.73%	56.84%
	Bigram	67.65%	46.44%	52.87%	48.49%
	Trigram	52.45%	76.40%	41.72%	36.31%
Support vector machine	Unigram	81.86%	77.60%	69.81%	70.68%
	Bigram	73.04%	83.98%	57.82%	54.79%
	Trigram	57.84%	80.67%	45.94%	41.86%
Logistic regression	Unigram	83.33%	79.38%	73.14%	74.54%
	Bigram	75.49%	51.81%	59.00%	54.63%
	Trigram	52.94%	40.94%	41.38%	35.56%

As shown in Table 2, the logistic regression model with unigram performs better than Support vector machine and Naïve Bayes. This model produces 83.33% accuracy, 79.38% precision, 73.14% recall, and 74.54% F-score. The model is then used to predict sentiment for the entire data. The result is as summarized in Table 3. From 19,783 reviews, 9,234 reviews (46.68%) had positive sentiments, 430 reviews (2.17%) had neutral sentiments, and 10,119 reviews (51.15%) had negative sentiments. Positive and negative sentiments do not differ much, around 4.47% or 885 reviews. However, reviews with negative sentiment still dominate, scoring 51.15%.

4.1.4. Topic modeling

After conducting sentiment analysis, topic modeling was carried out on 9,234 positive reviews (46.68%) to find out the topics that users liked and 10,119 negative reviews (51.15%) to find out the topics that users complained about. This study uses the latent dirichlet allocation algorithm for topic modeling and determines the number of topics based on the highest coherence score. In addition, the lambda values are explored to obtain relevant keywords, as in research [32]. For positive reviews, based on the experimental results, as shown in Figure 3, the highest coherence score obtained was 0.45, with a total of three topics. The topic modeling results for all positive review data are shown in Figure 3, with the best model achieved by using three topics. Topic naming is given based on the keywords extracted for each topic shown in Table 4. It is known that 36.40% discussed application features, 35.60% discussed products/services, and 28.10% discussed application interfaces.

Table 3. Sentiment classification results

Label	Total	Percentage
Positive	9,234	46.68%
Neutral	430	2.17%
Negative	10,119	51.15%

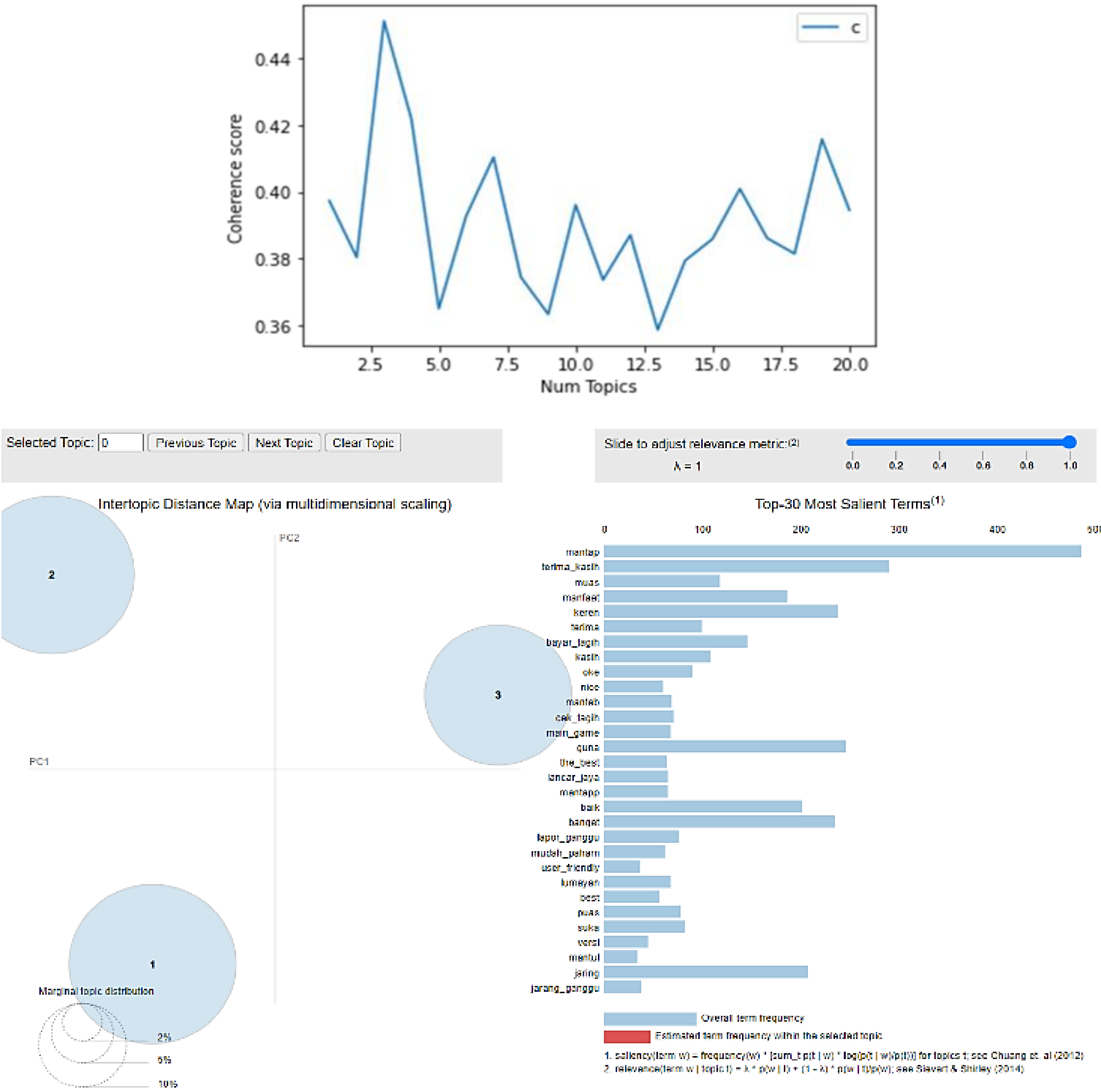

Figure 3. Coherence scores and topic cluster visualization for positive reviews

Table 4. Positive topics and keywords

No	Topic	Keyword	Percentage
1	Application features	*keren, manfaat, bayar_tagih, tampil, suka, fitur, lapor_ganggu, puas, lancar_jaya, mantap, lumayan, mudah_paham, lengkap, praktis, informasi, fitur_fiturnya, fiturnya, fiturnya_lengkap*	36.40%
2	Product/ Service	*terima_kasih, guna, baik, layan, main_game, the_best, pakai, respon, moga_depan, fast_respon, main, jarang_ganggu, cepat, adu, joss, informatif, jarang_kendala, tuker_poin, game, ramah, langgan_setia*	35.60%
3	Application interface	*mantap, muas, oke, mantab, cek_tagih, nice, cepat, lancar, versi, user_friendly, cs, cepat_tanggap, mantul, stabil, versi_baru, layan, add_on, friendly, transaksi*	28.10%

Meanwhile, for negative reviews, based on the experimental results, as shown in Figure 4, the highest coherence score obtained was 0.44, with a total of five topics. The topic modeling results for all negative review data are shown in Figure 4, with the best model achieved by using five topics. Topic naming is given based on the keywords extracted for each topic shown in Table 5. It is known that 22.80% discussed information availability, 21.80% discussed the reliability of application features, 19.50% discussed application processing speed, 18.90% discussed bugs, and 16.90% discussed application reliability.

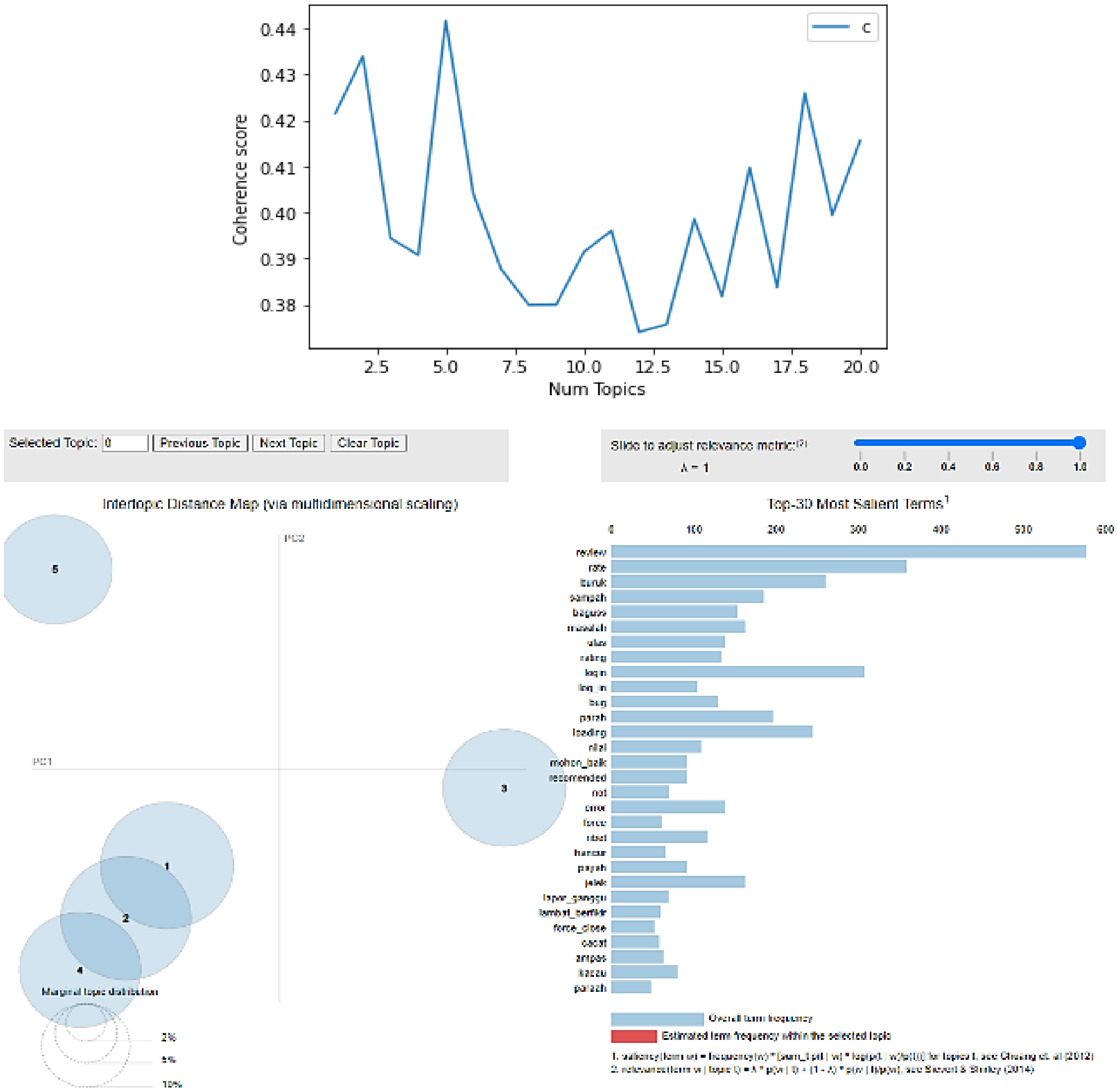

Figure 4. Coherence scores and topic cluster visualization for negative reviews

Table 5. Negative topics and keywords

No	Topic	Keyword	Percentage
1	Availability of information	*nilai, error, tagih, tampil, kacau, ampas, buka, versi_baru, responsif, response, versi, bayar_mahal, crash, riwayat_tagih, riwayat, bayar_tagih, hilang*	22.80%
2	Reliability of application features	*rate, login, recomended, eror, aneh, berat, gagal, habis_update, menu, loadingnya, susah_ampun, diupdate, no_hp, update, hp*	21.80%
3	Application processing speed	*review, buruk, bagus, loading, jelek, lapor_ganggu, ganggu, kecewa, fungsi, keluh, lapor, lama, super, lemot, cek_tagih, ram, lamban*	19.50%
4	Bugs	*sampah, ulas, rating, bug, payah, lambat_berfikir, user, masuk, akun, verif, berfikir, fix, please, user_friendly, susah, loading melulu, email, nomor_hp, nomor_telepon, parah, nge_lag, freeze*	18.90%
5	Application reliability	*masalah, parah, force, hancur, force_close, parah, ribet, lag, apknya, restart, lemot, close, restart, bikin_emosi, kode_verifikasi, sandi salah*	16.90%

4.2. Discussion

This study uses sentiment analysis and topic modeling to find out the service quality of a self-service mobile application in the telecommunications sector based on reviews from Google Play Store and Apple App Store. This study uses myIndiHome mobile application as a case study because it is one of the self-service mobile applications in the telecommunication sector with the most users in Indonesia. From this study, it is known that the logistic regression model with unigram gives better performance than support vector machine and Naïve Bayes as a whole. This model produces 83.33% accuracy, 79.38% precision, 73.14% recall, and 74.54% F-score. Meanwhile, support vector machine with unigram became the second classifier with the best performance, namely 81.86% accuracy, 77.60% precision, 69.81% recall, and 70.68% F-score. These results are similar to previous studies conducted by [22] and [33], where logistic regression and support vector machine provide accurate results that are similar. In addition to the algorithm used, the resulting level of accuracy is also influenced by the division of words used to create the model. As shown in Figure 5, models with unigram provide better accuracy values than models with bigram and trigram. This result is in line with previous research, Tiffani [34] shows Naïve Bayes with unigram produced the highest level of accuracy compared to Naïve Bayes with bigram and trigram. Shahana and Oman [35] also showed a similar thing, where in that study, support vector machine with unigram provided better accuracy than bigram.

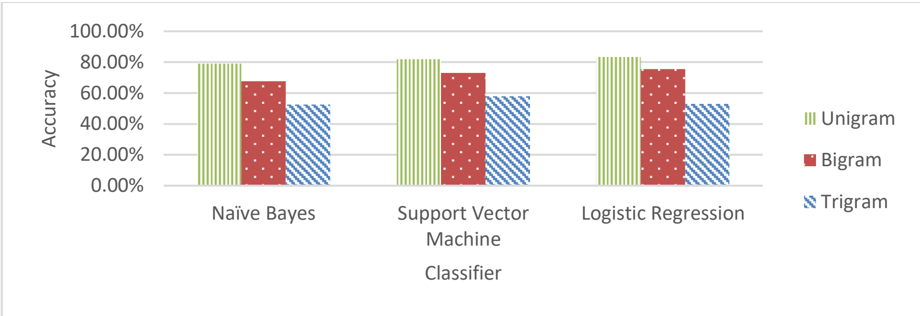

Figure 5. Comparison of accuracy of unigram, bigram, and trigram models

This study also produces sentiment classifications with positive, neutral, and negative sentiments for myIndiHome mobile application reviews from Google Play Store and Apple App Store, as shown in Table 3. Positive and negative sentiments have a difference that is not much different, around 4.47%. However, reviews with negative sentiment still dominate with a value of 51.15%, compared to positive sentiment of 46.68% and neutral sentiment of 2.17%. In other words, reviews of mobile self-service applications in the telecommunications sector are still dominated by negative sentiment, so improving the application based on these reviews is necessary. Figure 6 displays a word cloud related to myIndiHome application to see the words that appear most frequently in reviews. Words that appear most frequently are visualized as more significant in size and darker in color. Green words represent positive sentiments, red words represent negative sentiments, and blue words represent neutral sentiments. Word cloud visualization shows that reviews with positive sentiment relate to user satisfaction with applications that can help users. Moreover, reviews with negative sentiment relate to features in the application, while reviews with neutral sentiment relate to application ratings.

Figure 6. Wordcloud visualization

Sentiment classification results are further analyzed with topic modeling to find out what users like and opportunities for improvement in the application. Using latent dirichlet allocation and the number of topics based on the highest coherence score, three topics for positive reviews and five topics for negative reviews was obtained. In positive reviews, 36.40% discussed application features, 35.60% discussed products/services, and 28.10% discussed application interfaces. While in the negative reviews, 22.80% discussed information availability, 21.80% discussed the reliability of application features, 19.50% discussed application processing speed, 18.90% discussed bugs, and 16.90% discussed application reliability. The topics and keywords obtained in this negative review align with the factors influencing service satisfaction of self-service mobile applications in the telecommunications industry, including application speed, unwanted information, incomplete information, unavailable information/services, application response failure rate, and difficulty navigating [5]. The main discussion topics obtained are also related to the mobile application service quality dimensions described in Table 1 [18]. Next, the relationship between topics and MSAQ dimensions will be explained.

The first topic of negative sentiment reviews is the availability of information. Information is one of the things needed to be accessed by users from mobile applications. For this sector, information related to billing is essential. In negative sentiment reviews, many customers complain that billing information cannot be seen from the mobile application. Mobile application providers must provide and display the information users need [18]. It is also necessary to ensure that the information provided must be accurate and precise [18]. The topic of availability of information can be further categorized into the information dimension.

The second topic of negative sentiment reviews is the reliability of the application features. This topic concerns the reliability of a single feature on a mobile application. Features that cannot operate properly due to technical reasons are what users complain about. For example, complaints about difficult data updates causing failure when logging in to the mobile application were found in negative sentiment reviews. Damage to one feature can also affect other features and can even cause damage to the entire mobile application [18]. So mobile application providers must ensure that the features are executed according to the description and level of service promised [18]. The topic of application feature reliability can be further categorized into the technical reliability dimension.

The third topic on negative sentiment reviews is application processing speed. The topic of application processing speed refers to the processing performance of any operation in a mobile application, such as page loading, page transitions, and fast response to customer input. Speed of processing is also related to the mobile application's ability to display information quickly, such as visual graphics on mobile applications, and the quality of processing and data transfer [18]. While users usually cannot distinguish between application speed and network issues, it only means that all other apps generally work while certain apps are very slow [5]. In negative sentiment reviews, customers complain of a slow process when using any of the functions in the app. Mobile application providers must ensure that mobile applications have short waiting times and react responsively to customer interactions [18]. In [5], the quality and speed of application performance are important factors affecting the satisfaction level with mobile applications. The topic of application processing speed can be further categorized into the performance dimension.

The fourth topic on negative sentiment reviews is bugs. Bugs are related to problems in the application that must be fixed, such as crashes, wrong behavior, or performance problems [36]. Bugs need to be fixed through updates to the mobile application to improve the service quality perceived by customers and provide necessary services [18]. Bugs can be related to the ongoing operation to ensure mobile application updates are carried out. The topic of bugs can be further categorized into the technical reliability dimension.

The last topic on negative sentiment reviews is application reliability. This topic deals with the reliability of mobile applications in uninterrupted operations. For example, in negative sentiment reviews, complaints about force close, freeze, or lag issues were found. Such interruptions and malfunctions require restarting action, which involves the risk of losing previously acquired information, such as selected products or entered data. This causes inconvenience to the user. So mobile application providers must ensure that mobile applications run according to the description and level of service promised [18]. The topic of application feature reliability can be further categorized into the technical reliability dimension.

The discussion topics from the negative sentiment reviews are known to fall into information, performance, and technical reliability dimensions. The three secondary dimensions obtained each represent a different primary dimension. Information is part of the primary dimension of interaction quality, which shows the quality of interaction between customers and service providers [18]. Performance is part of the primary dimension of environment quality, which shows the context of mobile application delivery and the quality characteristics that affect the delivery of these mobile applications [18]. Then, technical reliability is part of the primary dimension of outcome quality, which shows the technical quality of service delivery and customer satisfaction with mobile services [18]. This shows that the application still needs improvement in various aspects to improve the quality of mobile application services. The main improvement is needed in outcome quality, where the three main topics in the negative sentiment review fall into technical reliability, feature reliability, application reliability, and bugs. This result is also in line with Bhale and Bedi [5], where most users make bugs, errors, and application response failures as reasons for dissatisfaction with the application, users do not care about technical problems when it comes to services in the application.

While on positive reviews, the first topic obtained is the application features. The topic of application features relates to overall satisfaction with mobile services through meeting customer needs and requirements on mobile applications. In positive sentiment reviews, many users are satisfied with the mobile application's features. This topic can be further categorized into the valence dimension, which describes customer satisfaction when using the application. The second topic obtained in positive reviews is the application interface. Application interface topics relate to visual aesthetics, layout clarity, and application ease of use. In positive sentiment reviews, users indicated their satisfaction with the application’s user-friendly interface. This topic can further classify into design dimensions that describe the aesthetics and layout of user interface designs. The last topic to get positive reviews is product/service. In positive sentiment reviews, users indicate their satisfaction with the availability of products and services that rarely experience interruptions and problems. This product/service topic cannot be mapped to the mobile application service quality dimensions because this topic is more focused on concrete services provided to customers. It relates more to the tangible SERVQUAL dimension to represent traditional services.

Topics of discussion from positive sentiment reviews are known to fall into valence and design dimensions. The two secondary dimensions obtained represent the primary dimensions of outcome quality and environment quality. This shows that several aspects of the primary dimensions of outcome quality and environment quality are considered good enough and need to be maintained in this mobile application.

4.3. Implication

For theoretical implications, this study uses sentiment analysis and topic modeling to find out the service quality of a self-service mobile application in the telecommunications sector from reviews on Google Play Store and Apple App Store. The service quality dimension refers to the mobile application service quality, which focuses more on mobile applications that run on mobile devices to provide mobile services to users [18]. Previous research [5] used a qualitative study based on a survey to measure customer engagement and satisfaction with service channels and the reasons for dissatisfaction with digital self-service.

Meanwhile, the practical implications of this study are opportunities for application improvement by analyzing complaints, needs, and input from users regarding mobile applications that are used so that mobile application providers, in general, and related organization, in particular, can create applications that meet the expectations of their users. This study can be input in developing applications that meet user expectations. In addition, implementing a system to monitor user reviews can also be carried out to get real-time input. If user complaints can be resolved, then user satisfaction with the application can increase.

To improve application quality, application providers can implement best coding practices in development. This includes writing clean code, checking for errors and vulnerabilities in the source code with regular code reviews and static code analysis, as well as testing automation. In addition, developing a mobile architecture that facilitates sustainability and scalability is also important to consider.

5. CONCLUSION

This study uses sentiment analysis and topic modeling to find out the service quality of a self-service mobile application in the telecommunications sector based on reviews from Google Play Store and Apple App Store. myIndiHome mobile application is used as a case study because it is one of the self-service mobile applications in the telecommunication sector with the most users in Indonesia. This study shows that the logistic regression model with unigram gives the best performance compared to the support vector machine and Naïve Bayes model, with 83.33% accuracy, 79.38% precision, 73.14% recall, and 74.54% F-score. The level of accuracy produced is also influenced by word division. The model with unigram gives a better accuracy value than the model with bigram and trigram. The sentiment analysis results show that negative sentiment dominates with a score of 51.15%, compared to 46.68% positive sentiment and 2.17% neutral sentiment. The topic modeling results show that the positive review data has three topics, including application features, products/services, and application interfaces. While the negative review data has five topics, including application availability, application feature reliability, application processing speed, bugs, and application reliability. The discussion topics from the negative sentiment reviews are mapped into information, performance, and technical reliability dimensions. It shows that the application still needs improvement in various aspects. Moreover, the discussion topics from the positive sentiment reviews are mapped into the valence and design dimensions category. It shows that several aspects of the primary dimension of outcome quality and environment quality are considered good enough and must be maintained. Based on this research, the recommendation for mobile application providers is to minimize user complaints regarding discussion topics from negative sentiment reviews to increase user satisfaction. A suggestion for future research is to validate customer satisfaction levels with mobile self-service applications in other telecommunications sectors. Besides that, it can use other algorithms to provide better performance.

REFERENCES

[1] Data.ai, "State of mobile 2022 Indonesia," *Data.Ai*, 2022, [Online]. Available: https://www.data.ai/en/go/state-of-mobile-2022-indonesia/.

[2] Statista Research Department, "Number of smartphone subscriptions worldwide from 2016 to 2021, with forecasts from 2022 to 2027 ," *Statista*, vol. 2027, pp. 2–5, 2022, [Online]. Available: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.

[3] L. Ceci, "App stores statistics," *Statista*, 2022, [Online]. Available: https://www.statista.com/topics/1729/app-stores/.

[4] Statista, "Mobile app usage - statistics & facts," *Statista*, pp. 1–16, 2019.

[5] U. A. Bhale and H. S. Bedi, "A qualitative study on service channels in the indian telecom industry," *International Journal of Scientific and Technology Research*, vol. 9, no. 3, pp. 265–270, 2020, [Online]. Available: www.ijstr.org.

[6] S. Dhar and I. Bose, "Walking on air or hopping mad? understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps," *Decision Support Systems*, vol. 162, 2022, doi: 10.1016/j.dss.2022.113769.

[7] N. Genc-Nayebi and A. Abran, "A systematic literature review: opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, pp. 207–219, 2017, doi: 10.1016/j.jss.2016.11.027.

[8] B.-H. Leem and S.-W. Eum, "Using text mining to measure mobile banking service quality," *Industrial Management & Data Systems*, vol. 121, no. 5, pp. 993–1007, Apr. 2021, doi: 10.1108/IMDS-09-2020-0545.

[9] O. Oyebode, F. Alqahtani, and R. Orji, "Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews," *IEEE Access*, vol. 8, pp. 111141–111158, 2020, doi: 10.1109/ACCESS.2020.3002176.

[10] M. Nayebi, H. Cho, and G. Ruhe, "App store mining is not enough for app improvement," *Empirical Software Engineering*, vol. 23, no. 5, pp. 2764–2794, Oct. 2018, doi: 10.1007/s10664-018-9601-1.

[11] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: a survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: 10.1016/j.asej.2014.04.011.

[12] A. R. Pathak, M. Pandey, and S. Rautaray, "Topic-level sentiment analysis of social media data using deep learning," *Applied Soft Computing*, vol. 108, 2021, doi: 10.1016/j.asoc.2021.107440.

[13] V. Balakrishnan, Z. Shi, C. L. Law, R. Lim, L. L. Teh, and Y. Fan, "A deep learning approach in predicting products' sentiment ratings: a comparative analysis," *The Journal of Supercomputing*, vol. 78, no. 5, pp. 7206–7226, Apr. 2022, doi: 10.1007/s11227-021-04169-6.

[14] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[15] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: approaches, challenges and trends," *Knowledge-Based Systems*, vol. 226, 2021, doi: 10.1016/j.knosys.2021.107134.

[16] P. Kherwa and P. Bansal, "Topic modeling: a comprehensive review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 7, no. 24, pp. 1–16, 2020, doi: 10.4108/eai.13-7-2018.159623.

[17] I. Vayansky and S. A. P. Kumar, "A review of topic modeling methods," *Information Systems*, vol. 94, 2020, doi: 10.1016/j.is.2020.101582.

[18] T. Wulfert, "Mobile app service quality dimensions and requirements for mobile shopping companion apps," *Junior Management Science*, vol. 4, no. 3, pp. 339–391, 2019, doi: 10.5282/jums/v4i3pp339-391.

[19] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality," *Journal of retailing*, vol. 64, no. 1, pp. 12–40, 1988.

[20] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "E-S-QUAL a multiple-item scale for assessing electronic service quality," *Journal of Service Research*, vol. 7, no. 3, pp. 213–233, 2005, doi: 10.1177/1094670504271156.

[21] E. Y. Huang, S. W. Lin, and Y. C. Fan, "M-S-QUAL: mobile service quality measurement," *Electronic Commerce Research and Applications*, vol. 14, no. 2, pp. 126–142, 2015, doi: 10.1016/j.elerap.2015.01.003.

[22] B. Liu, "Sentiment analysis: mining opinions, sentiments, and emotions," *Sentiment Analysis*, 2020, doi: 10.1017/9781108639286.

[23] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," 2010, pp. 223–239.

[24] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Systems with Applications*, vol. 57, pp. 117–126, 2016, doi: 10.1016/j.eswa.2016.03.028.

[25] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012, doi: 10.1145/2133806.2133826.

[26] D. M. Blei, A. Y. Ng, and M. T. Jordan, "Latent dirichlet allocation," *Advances in Neural Information Processing Systems*, vol. 3, pp. 993–1022, 2002.

[27] G. P. T. Indonesia, "myIndiHome," 2016, [Online]. Available: https://play.google.com/store/apps/details?id=com.telkom.indihome.external.

[28] A. S. T. Indonesia, "myIndiHome," 2016, [Online]. Available: https://apps.apple.com/id/app/myindihome/id1119407221.

[29] X. Hu and H. Liu, "Text analytics in social media," *Mining Text Data*, vol. 9781461432, pp. 385–414, 2012, doi: 10.1007/978-1-4614-3223-4_12.

[30] S. García, J. Luengo, and F. Herrera, "Introduction," *Intelligent Systems Reference Library*, vol. 72, pp. 1–17, 2015, doi: 10.1007/978-3-319-10247-4_1.

[31] G. Miner, J. Elder, A. Fast, T. Hill, R. Nisbet, and D. Delen, "Practical text mining and statistical analysis for non-structured text data applications," *2012*.

[32] C. Sievert and K. Shirley, "LDAvis: a method for visualizing and interpreting topics," pp. 63–70, 2015, doi: 10.3115/v1/w14-3110.

[33] B. Andrian, T. Simanungkalit, I. Budi, and A. F. Wicaksono, "Sentiment analysis on customer satisfaction of digital banking in Indonesia," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 3, pp. 466–473, 2022, doi: 10.14569/IJACSA.2022.0130356.

[34] I. E. Tiffani, "Optimization of Naïve Bayes classifier by implemented unigram, bigram, trigram for sentiment analysis of hotel review," *Journal of Soft Computing Exploration*, vol. 1, no. 1, 2020, doi: 10.52465/joscex.v1i1.4.

[35] P. H. Shahana and B. Omman, "Evaluation of features on sentimental analysis," *Procedia Computer Science*, vol. 46, pp. 1585–1592, 2015, doi: 10.1016/j.procs.2015.02.088.

[36] W. Maalej and H. Nabil, "Bug report, feature request, or simply praise? on automatically classifying app reviews," in *2015 IEEE 23rd International Requirements Engineering Conference, RE 2015 - Proceedings*, 2015, pp. 116–125, doi: 10.1109/RE.2015.7320414.

BIOGRAPHIES OF AUTHORS

Jwalita Galuh Garini is currently a master's student in information technology at the Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia. She received the bachelor's degree in information systems from Sepuluh Nopember Institute of Technology, Surabaya in 2018. She is interested in data mining and machine learning. She can be contacted via email: jwalita.galuh@ui.ac.id or taajwalita@gmail.com.

Achmad Nizar Hidayanto is a professor of information systems at Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia. He received the bachelor's, master's, and Ph.D. degrees in computer science from University of Indonesia, Jakarta in 1999, 2002, and 2008. His research interests are related to information systems, information technology adoption, digital economy, and financial technology. He can be contacted at email: nizar@cs.ui.ac.id.

Agri Fina is currently a master's student in information technology at the Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia. She received the bachelor's degree in information systems from Del Institute of Technology, Toba in 2018. She is interested in UI/UX design, product design, graphic design, human-computer interaction, and machine learning. She can be contacted via email: agri.fina@ui.ac.id.