Analysis of Spotify User Sentiment to Improve Customer Satisfaction Using Opinion Mining and Latent Dirichlet Allocation Based on E-Satisfaction Dimensions

[1]Mutawakkil Samjas, [2]Armin Darmawan
[1,2]Department of Industrial Engineering, Hasanuddin University, Makassar, Indonesia
Email: [1]mutawakkilsamjas88@gmail.com, [2]darmawanarmin@gmail.com

Article Info
Article history:
Received Oct 10th, 2025
Revised Nov 02nd, 2025
Accepted Nov 30th, 2025

Keyword:
E-Satisfaction
Latent Dirichlet Allocation
Naïve Bayes
Sentiment Analysis
Support Vector Machine

ABSTRACT
This study aims to enhance Spotify customer satisfaction by analyzing user reviews on the Google Play Store using sentiment analysis techniques and identifying relevant topics related to customer satisfaction based on the dimensions of electronic satisfaction. The methods used in this analysis are Support Vector Machine (SVM), Naïve Bayes (NB), and Latent Dirichlet Allocation (LDA). The results show that SVM is the most effective technique for text classification, with accuracies of 87%, 87%, 81%, and 84%, respectively, along with precision, recall, and F1-score of 0.93, 0.93, and 0.84. LDA was utilized to extract various topics within the e-satisfaction dimensions, with serviceability emerging as the top priority for improvement. Identified topics include connectivity and accessibility, performance and user experience, premium services, app quality, content and playlists, app features, and sound/music quality. These findings suggest that improvements in server infrastructure, the implementation of AI-driven chat support, enhanced ad management, and improved song lyrics databases could substantially enhance Spotify's customer satisfaction.

Corresponding Author:
Mutawakkil Samjas
Department of Industrial Engineering,
Hasanuddin University,
Jl. Perintis Kemerdekaan Km. 10, Tamalanrea, Makassar 90245, Sulawesi Selatan, Indonesia.
Email: mutawakkilsamjas88@gmail.com
DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.38480

1. INTRODUCTION

The increase in internet users is in line with the rapid growth of smartphone adoption nationally. Smartphone users in 2024 reached 201.97 million [1]. According to the Ministry of Communication and Information, around 167 million people, equivalent to 89% of the total population, are smartphone users [2]. Since the beginning of the pandemic, smartphone usage by the public has changed. In 2023, the average Indonesian spent around 6.05 hours daily on their mobile devices, according to The State of Mobile 2024 [3]. This shift reflects a profound change in societal behavior, where digital devices are increasingly relied upon for communication, work, and entertainment. The widespread adoption of digital technologies has become a central element of daily routines, influencing how people live, communicate, and interact with each other.

One sector notably impacted by digital transformation is the music industry. Music, as a medium of emotional expression and cross-cultural communication, is enjoyed by people from all walks of life. Today, most Indonesians prefer to access music via smartphone applications, which offer greater convenience compared to traditional media such as CDs, DVDs, or radio [4]. A survey conducted by the Indonesian Internet Service Providers Association (APJII) in 2016 reported that 46.9 million Indonesians listened to music online. Furthermore, according to APIJI, several music applications emerged as the most frequently used platforms by Indonesian users in 2023, as illustrated in Figure 1 [5].

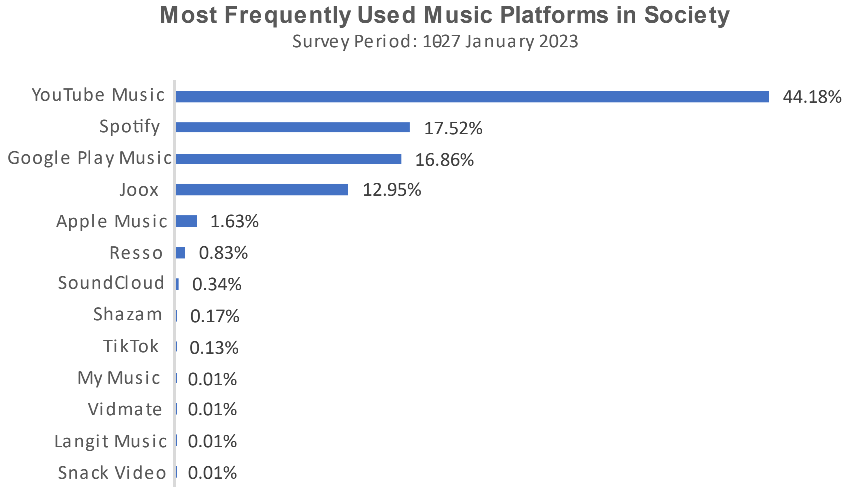


Figure 1. Most used music apps in Indonesia [2]

Figure 1 shows that Indonesians tend to prefer YouTube Music (44.18%), while Spotify only accounts for 17.52%. This gap indicates that YouTube Music has successfully attracted users through its superior features. Intense competition is also taking place between Spotify and other apps like Google Play Music. This situation demands that Spotify continuously innovate and improve its service quality, in line with Norman's (2016) statement that the main factors influencing user loyalty to an app are user perception or experience, a good relationship with the app, and satisfaction with the services provided [6].

Although Spotify ranks second as the leading music app within the Indonesian market, the decline in its rating on the Google Play Store, from 4.4 in 2023 to 4.2 in 2024, suggests a growing level of customer dissatisfaction [7]. This decrease indicates that Spotify has not entirely fulfilled customer expectations, presenting a significant challenge for the company in maintaining and improving its position in the competitive market.

Customer satisfaction is closely related to user loyalty, where positive experiences increase the likelihood of customers continuing to use the app [8]. Therefore, to improve customer satisfaction, it is important to analyze the sentiments contained in user comments, such as those found on platforms like Google Play Store. User comments, whether positive or negative, reflect user perceptions of Spotify, which are crucial indicators for the company to make improvements. In a competitive and dynamic market, customer comments directly impact the success of a product, as they reflect the feelings and expectations of customers toward the product they use [9].

As demonstrated by the research conducted by Chong and Shah (2022), Rahman and Seddiqui (2019), and Guia et al. (2019), sentiment analysis can be carried out using techniques such as Support Vector Machine (SVM) and Naïve Bayes (NB) [10][11][12]. The key difference between this study and previous research lies in the research object being analyzed, the data sources used, and the inclusion of an additional method, Latent Dirichlet Allocation (LDA).

LDA enables topic modeling, providing a better understanding of customer concerns. This method helps identify areas that require improvement by focusing on various dimensions of e-satisfaction, including convenience, site design, usability, merchandising, and security [13]. Therefore, this study aims to offer deeper insights into Spotify's customer satisfaction and provide strategic recommendations for the company to enhance customer satisfaction and strengthen its market position.

2. RESEARCH METHOD

2.1 Research Flowchart

This study uses sentiment analysis to measure user satisfaction with the Spotify app based on comments collected from the Google Play Store between February 20, 2024, and February 20, 2025. In this study, two sentiment analysis techniques are utilized is SVM and NB. The primary objective of this analysis is to assess the performance of each algorithm in accurately classifying user sentiments, as well as to gain deeper insights into user satisfaction with the app. Additionally, to enhance the understanding of user satisfaction, this research also employs LDA method to identify the main topics frequently discussed by users, particularly those related to e-satisfaction dimensions that received the highest negative responses. The research process can be seen in Figure 2.

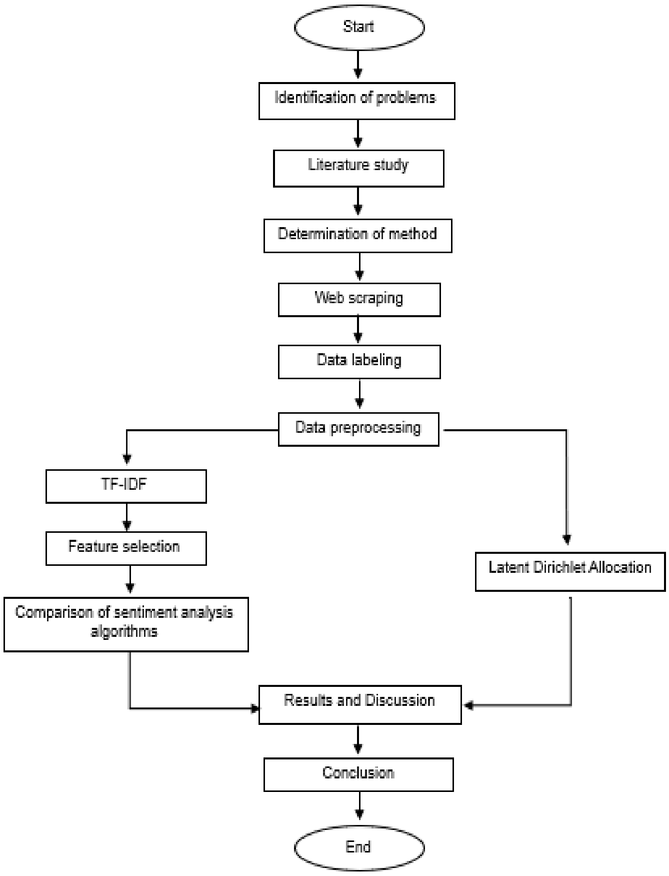


Figure 2 Flowchart of the Research

Based on Figure 2, the research process begins with problem identification, which is the first step in determining the issues faced and the research objectives, as outlined in the background. After that, a literature review is conducted to gather relevant theories, including theories on e-satisfaction, sentiment analysis, topic modeling, and model evaluation, which will be used in this study. Further discussion of these theories will be presented in the next section.

The user comments data for Spotify used sin this study is gathered using a web scraping technique from user reviews on the Google Play Store through Python programming on the Google Colab platform. This technique allows for the efficient collection of large volumes of relevant data for subsequent analysis. Once the data is collected, a manual labeling process is applied, where each user review is categorized based on sentiment (positive or negative) and e-satisfaction (convenience, site design, serviceability, merchandising, and security). This labeling process ensures that the data used in the analysis is both accurate and pertinent to the research objectives.

After the data has been labeled, the subsequent step involves data preprocessing. This phase includes cleaning the data by removing unnecessary symbols, numbers, and emojis, transforming the text into lowercase through case folding, breaking the text into words or tokens through tokenization, removing stopwords that carry little meaning, and applying stemming techniques to simplify words to their base forms. The goal of the preprocessing stage is to prepare the data for more detailed analysis. Following preprocessing, feature extraction is carried out using the TF-IDF method, which identifies the most important words in each comment. Next, feature selection is applied to eliminate unnecessary features, ensuring that only the most significant ones are retained for the analysis.

Sentiment analysis is carried out using two algorithms. The SVM algorithm is employed to determine the optimal boundary that divides sentiment categories, while NB calculates probabilities to classify the sentiment in user comments. Additionally, topic modeling is conducted using the LDA approach, which is applied to uncover the primary topics frequently discussed by users in their feedback.

The results of the sentiment analysis and topic modeling are then presented within the section on results and discussion, where the key insights of this study will be explored, and recommendations for enhancing Spotify user satisfaction will be provided. The conclusion will offer a summary of the outcomes of this research.

2.2 Literature Review

2.2.1 E-Satisfaction

According to Komara (2014), e-satisfaction is defined as the satisfaction experienced after making a purchase in an industry through electronic services [14]. Ranjbarian et al. (2012) identified five main dimensions that influence e-satisfaction, which include:

1. Convenience refers to how users perceive comfort and ease when interacting with online services. This includes the ability of customers to navigate features that align with their preferences, as well as the time-saving aspects and overall comfort experienced during the use of online platforms.
2. Merchandising refers to the availability of products, the variety of choices, and the appeal of the products or services offered, which can enhance the likelihood of fulfillment and user satisfaction.
3. Site design refers to the quality of the design and layout of the website or application, with a simple navigation flow, quick presentation, and ease of navigation that saves the user's time.
4. Security refers to the protection of personal data within the application, which is a crucial factor in preventing potential fraud.
5. Serviceability refers to the extent to which the application can provide stable services with minimal downtime or errors, manage regularly updated content, offer responsive customer support, perform routine feature upgrades, and manage relevant advertisements without disrupting the user experience [13].

2.2.2 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF technique is used to assess the significance of a term within a specific document in relation to the entire set of documents. The formulas for TF and IDF are presented in equations 1, 2, and 3 [15].

$$\mathrm{TF}(w_i) = \frac{Nw_i}{Nw} \tag{1}$$

$$\mathrm{IDF}(w_i) = \log\left(\frac{Nd}{Nd/w_i}\right) \tag{2}$$

$$\text{TF-IDF}(w_i) = \mathrm{TF}(w_i) \text{ x } \mathrm{IDF}(w_i) \tag{3}$$

2.2.3 Feature Selection

Feature selection is a process that aims to reduce the dimensionality of the feature vector by selecting the most significant attributes. A widely used approach for feature analysis is the Chi-Square algorithm. Using the Chi-Square test can significantly reduce the feature set, thereby enhancing classification performance. The Chi-Square formula is shown in Equation 4 [16].

$$\text{Chi-Square } (t_k, c_i) = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \tag{4}$$

2.2.4 Sentiment Analysis Algorithms

According to Thelwall, as cited in Haddi, sentiment analysis is considered a process of categorizing texts to determine if their sentiment is positive or negative [17]. In this research, the classification methods applied for sentiment analysis are:

1. Support Vector Machine (SVM)
   SVM is widely regarded as an effective method for text classification, operating by identifying the optimal hyperplane or dividing line that differentiates between two distinct classes. To find the best hyperplane, SVM searches for the separating line that lies exactly in the middle between the two classes, ensuring that the margin between them is maximized. This process helps improve the separation between classes, leading to more optimal classification performance [18]. The determination of the SVM hyperplane is formulated using the following equation 5-7 [19]:

$$(w.x_i) + b = 0 \tag{5}$$

$$(w.x_i) + b \leq 1,\ y_i = -1 \tag{6}$$

$$(w.x_i) + b \leq 1,\ y_i = 1 \tag{7}$$

2. Naïve Bayes (NB)
   NB is a popular method in text mining, frequently used for text classification by applying probability and statistical techniques developed by Thomas Bayes. This method is highly competent and performs well in classification tasks. The NB formula is presented in Equation 8 [17]:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \tag{8}$$

2.2.5 Latent Dirichlet Allocation (LDA)

LDA is a probabilistic method used to study and represent a set of documents, with the goal of identifying latent topics within them. This method applies a probabilistic model to analyze the documents, determining a range of possible topics. LDA views each document as a combination of multiple topics, with each topic linked to a probability distribution across the terms that characterize it [20]. The LDA formulation 9 [21]:

$$P(W,Z,\theta,\varphi,\alpha,\beta) = \prod_{j=1}^{M} P(\theta_j;\alpha) \prod_{i=1}^{K} P(\varphi_i;\beta) \prod_{t=1}^{N} P\left(Z_{j,t}\middle|\theta_j\right) P\left(W_{j,t}\middle|\varphi_{j,t}\right) \tag{9}$$

2.2.6 Model Evaluation

Evaluating the model is a crucial process for determining its accuracy and reliability. In this research, three types of model evaluation are applied: the use of the confusion matrix and the Receiver Operating Characteristic (ROC) curve for evaluating the sentiment analysis model, and the coherence score for assessing the LDA model.

1. Confusion matrix (see Table 1)

Table 1. Confusion matrix

Actual	Predicted	
	+	-
+	True Positive	False Positive
-	False Negative	True Negative

A confusion matrix serves as a widely utilized instrument to evaluate the effectiveness of a classification algorithm. It includes four crucial values necessary for assessing the performance of the classifier: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The results from the confusion matrix are used to calculate several performance metrics, including accuracy, precision, recall, and F1-score. These calculations help evaluate the model's ability to classify data effectively. The formulas for these metrics are provided in [22].

a. Accuracy indicates the proportion of correct classifications overall. This is determined using Equation 10 as follows:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100\% \tag{10}$$

b. Precision measures how many positive predictions are correct. It is calculated using Equation 11 as follows:

$$\text{Precision} = \frac{(TP)}{(TP+FP)} \times 100\% \tag{11}$$

c. Recall measures how many actual positives were correctly predicted. It is calculated using Equation 12 as follows:

$$\text{Recall} = \frac{(TP)}{(TP+FN)} \times 100\% \tag{12}$$

d. The F1-score is the harmonic average that balances both precision and recall. It is determined using Equation 13 as shown below:

$$\text{F1-score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \times 100\% \tag{13}$$

2. Curve ROC
   ROC is a technique used to assess the performance of binary classification models at different decision thresholds. The ROC curve shows the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) at various decision points [23]. A critical measure obtained from the ROC is the Area Under the Curve (AUC), which quantifies the area beneath the ROC curve. Table 2 is the classification for AUC values.

Table 2. AUC values

Area under the curve (AUC)	Interpretation
$0.9 \leq AUC$	Excellent
$0.8 \leq AUC < 0.9$	Good
$0.7 \leq AUC < 0.8$	Fair
$0.6 \leq AUC < 0.7$	Poor
$0.5 \leq AUC < 0.6$	Fail

3. Coherence score
   Coherence in LDA plays a vital role in evaluating the relevance and comprehensibility of the topics produced by the model. Common coherence measures like UMass, UCI, and C_v assess the connections between words in a topic through various methods [24]. The C_v score is widely used due to its strong correlation with human evaluations of topic clarity. A higher coherence score signifies improved quality of the topics [25]. Additionally, this score helps determine the optimal number of topics in the LDA model by selecting the topic count that yields the highest coherence score.

3. RESULTS AND ANALYSIS

This study successfully collected 13,245 user comments from user reviews of the Spotify app on the Google Play Store, gathered through web scraping methods. This process was guided by the keyword “com.spotify.music” to ensure relevance to the application context. All retrieved comments were written in Indonesian and underwent verification to confirm their contextual appropriateness. Table 3 presents selected examples of the comments used in the analytical process.

Table 3. Web scraping results

Date	Comment
2025-02-20 17:00:37	Spotify semakin lama semakin aneh, bukannya bikin buat pengguna nyaman malah bikin males karna banyak iklannya, dan pembatasannya itu astaga kek maksa banget disuruh beli premium, tolong kembalikan spotify yang dulu, maaf bintang 1 karna sangat kecewa dengan spotify yang sekarang 🙏

3.1 Data Labeling

The labeling process serves as a critical foundation for enabling algorithms to effectively interpret the collected data. Each retrieved comment was manually labeled to ensure accuracy and relevance for subsequent analysis. Sentiment labeling was conducted by categorizing the comments into two classes: positive and negative. In addition, the comments were also labeled according to e-satisfaction dimensions, which include the categories of convenience, merchandising, site design, security, and serviceability. Table 4 shows the data labeling results of this research.

Table 4. Data labeling results

Date	Comment	Sentiment	Dimension
2025-02-20 17:00:37	Spotify semakin lama semakin aneh, bukannya bikin buat pengguna nyaman malah bikin males karna banyak iklannya, dan pembatasannya itu astaga kek maksa banget disuruh beli premium, tolong kembalikan spotify yang dulu, maaf bintang 1 karna sangat kecewa dengan spotify yang sekarang 🙏	Negative	Convenience

3.2 Data Preprocessing

Preprocessing the data is an essential phase in preparing the dataset, as it has a major impact on the performance of classification models [26]. The preprocessing procedure involves a series of tasks, including cleaning, case folding, tokenizing, normalization, stopword removal, and stemming. Table 5 presents a comparison of the data before and after the cleaning process, which involved removing irrelevant elements such as emojis, mentions, URLs, hashtags, symbols, duplicate entries, and empty records.

Table 5. Data cleaning results

Type of Dataset	Dimension
Before cleaning	13,245
After cleaning	11,362

Table 6 presents selected examples of the data preprocessing results, illustrating the steps of case folding, tokenizing, normalization, stopword removal, and stemming.

Table 6. Data preprocessing results

Dikit² premium dikit² premium KEMBALIKAN SPOTIFY YANG DULU meski iklan seenggaknya bebas skip lagu	
Case folding	dikit² premium dikit² premium kembalikan spotify yang dulu
Tokenizing	['dikit²', 'premium', 'dikit²', 'premium', 'kembalikan', 'spotify', 'yang', 'dulu']
Normalization	['sedikit', 'premium', 'sedikit', 'premium', 'kembalikan', 'spotify', 'yang', 'dulu']
Stopwords	['premium', 'premium', 'kembalikan', 'spotify']
Stemming	['premium', 'premium', 'kembali', 'spotify']

3.3 Sentiment Analysis

Sentiment analysis is the process of identifying the emotional tone or perspective expressed in a piece of text, such as whether the sentiment is positive or negative. Before carrying out sentiment classification using the SVM and NB algorithms, several preparatory actions must be taken. These actions involve splitting the dataset into training and testing sets, applying TF-IDF, and performing feature selection, all of which are essential for enhancing the sentiment classification task.

1. Training and testing dataset split
   Splitting the dataset into distinct training and testing subsets is an essential step in building sentiment analysis models. This division is essential to prevent overfitting, optimize hyperparameters, and evaluate the model's predictive performance [27]. The dataset was split using a ratio of 70% for training data and 30% for testing data [27][28]. Table 7 presents the results of the dataset split into training and testing sets.

Table 7. Results of the dataset split

Type of Dataset	Number of Records	Proportion (%)
Training data	7,953	70
Testing data	3,409	30
Total	11,362	100

2. TF-IDF
   TF-IDF calculates word relevance by giving greater importance to terms that occur frequently in a particular document but are less common across the entire document set, thereby highlighting more informative and contextually significant words [29]. Table 8 shows the TF-IDF results for the top 3 words with the highest average TF-IDF scores.

Table 8. TF-IDF results

Word	TF-IDF score
lagu	0.081376
bagus	0.054267
iklan	0.048880

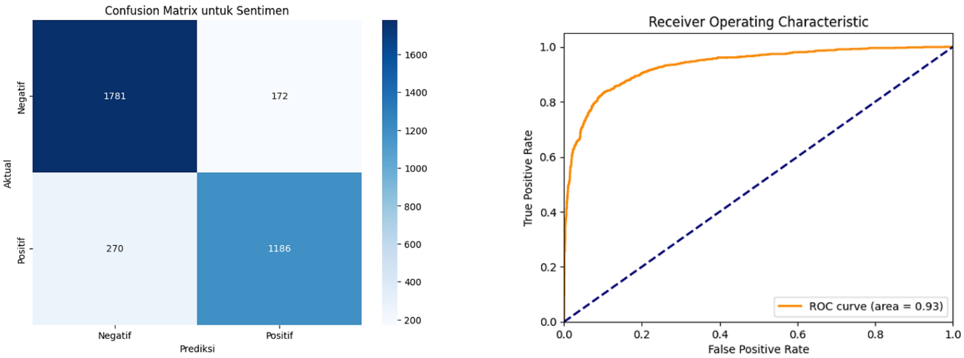
3. Feature selection
   The Chi-Square method is frequently used in feature selection during data processing to examine the relationship between features (such as words) and the target variable (sentiment). This approach evaluates the contribution of each feature to predicting the target class, especially when using categorical data [30]. Table 9 shows the ten words with the highest Chi-Square values.

Table 9. Feature selection result

Word	Chi square score
premium	156.770299
bagus	154.940818
enak	95.791684

4. Support Vector Machine (SVM)
   SVM is a popular classification technique that works by creating an optimal hyperplane to divide data into separate categories, aiming to maximize the margin between them. The selection of

parameters plays a critical role in determining the performance of SVM, thus requiring experimentation to identify the most effective configuration [31]. Based on the results of GridSearchCV, the optimal parameters obtained were C = 100, gamma = 0.01, and kernel = 'RBF'. To fully assess the algorithm's performance, evaluations were conducted using a confusion matrix and an ROC curve, as illustrated in Figure 3.

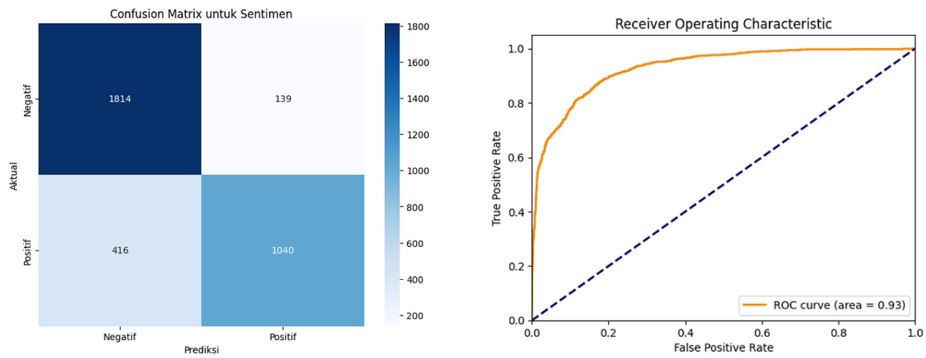


(a) Confusion matrix SVM (b) ROC curve SVM

Figure 3. Evalutions SVM

Referring to Figure 3(a), the values from the confusion matrix are used to calculate the evaluation metrics, including accuracy, precision, recall, and F1 score, as described in Equations 10–13. They yield 87%, 87%, 81%, and 84%, respectively. The ROC curve (Figure 3(b)) yields an AUC score of 0.93, indicating that the SVM algorithm performs very well in distinguishing between classes. This indicates that SVM can effectively distinguish between positive and negative classes with minimal error [32].

5. Naive Bayes (NB)
NB is a probabilistic classification algorithm that utilizes Bayes' theorem and conditional probability to make predictions based on observed features or words. It applies a statistical approach to inductive inference by calculating the probability of class labels given the observed data. NB is widely recognized for its speed, simplicity, and robustness against outliers [33]. One commonly used variant is Multinomial Naive Bayes, which is well-suited for frequency-based data such as text. After the NB algorithm is developed, a thorough evaluation is crucial to measure its performance, utilizing both a confusion matrix and an ROC curve. The results of the NB algorithm evaluation are shown in Figure 4.



(a) Confusion matrix NB (b) ROC curve NB

Figure 4. Evalutions NB

Referring to Figure 4(a), the values from the confusion matrix are used to calculate the evaluation metrics, including accuracy, precision, recall, and F1 score, as described in Equations 10–13. They yield 84%, 88%, 71%, and 78%, respectively. The ROC curve (Figure 4(b)) yields an AUC score of 0.93, indicating that the NB algorithm performs very well in distinguishing between classes. This indicates that NB can effectively distinguish between positive and negative classes with minimal error [32].

6. Comparison of SVM and NB Algorithms
   Determining which classification algorithm, SVM or NB, performs best can be achieved through a comprehensive evaluation based on several performance metrics. Table 10 presents a comparative summary of the performance of both algorithms.

Table 10. Comparative Results of SVM and NB Algorithms

Evaluation Metric	SVM	NB
Accuracy	87%	84%
Precision	87%	88%
Recall	81%	71%
F1-score	84%	78%
AUC	0.93	0.93
Duration	1218,457 s	0,112 s

   Based on Table 10, the SVM algorithm demonstrates superior performance compared to NB across nearly all evaluated metrics, although it requires a longer duration to build the model.

7. Dataset Analysis Using the Superior Algorithm

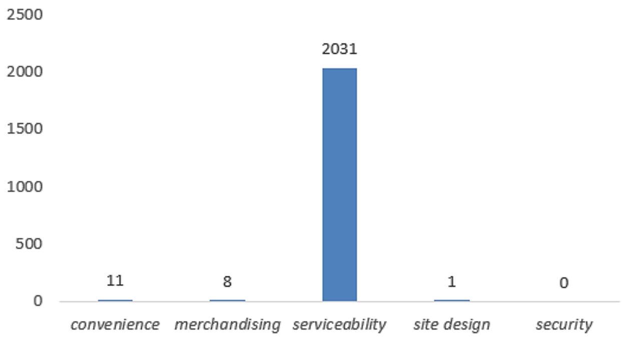


Figure 5. Negatively perceived e-satisfaction dimensions

   Following the comparative evaluation of the two algorithms, SVM was identified as the most effective, and was used to analyze negative user comments based on dimensions. The following insights were obtained can be seen in Figure 5.

3.4 Latent Dirichlet Allocation (LDA)

LDA is a commonly applied technique for topic modeling in text analysis, designed to uncover hidden thematic structures within large document collections. In this study, LDA was used to examine the serviceability dimension. The optimal number of topics was determined by selecting the one with the highest coherence score. As shown in Figure 6, the best number of topics is 16, yielding a coherence score of 0.482750.

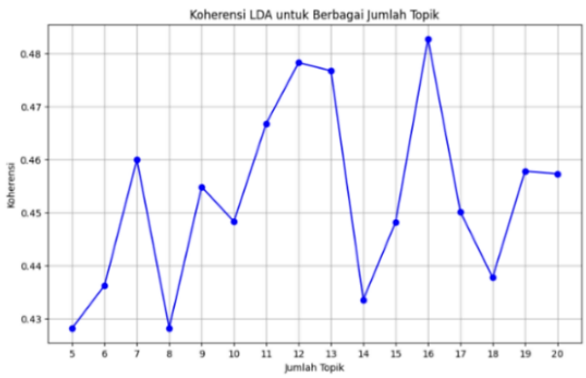


Figure 6. Coherence score results

After determining the optimal number of topics, the LDA model was retrained using this optimal value, which amounted to 16 topics. The results of this training are shown in Table 11.

Table 11. Topics within the serviceability dimension

Topic	Conten
Topic 1	0.071*"pakai" + 0.065*"offline" + 0.061*"spotify" + 0.040*"akun" + 0.032*"buka" + 0.031*"data" + 0.029*"aplikasi" + 0.029*"masuk" + 0.029*"wifi" + 0.027*"tolong"
Topic 2	0.110*"update" + 0.074*"seru" + 0.057*"kualitas" + 0.045*"horror" + 0.041*"lambat" + 0.036*"sulit" + 0.029*"audio" + 0.027*"tinggal" + 0.025*"sdh" + 0.022*"mengganggu"
Topic 3	0.152*"premium" + 0.090*"beli" + 0.040*"coba" + 0.038*"nya" + 0.035*"tolong" + 0.030*"bayar" + 0.029*"baik" + 0.026*"paket" + 0.025*"bug" + 0.023*"min"
Topic 4	0.065*"bagus" + 0.055*"bayar" + 0.055*"jaring" + 0.053*"aplikasi" + 0.041*"main" + 0.038*"kuota" + 0.037*"download" + 0.032*"game" + 0.029*"sinyal" + 0.028*"internet"
Topic 5	0.063*"music" + 0.040*"lancar" + 0.031*"parah" + 0.025*"minggu" + 0.023*"baik" + 0.023*"full" + 0.021*"the" + 0.020*"langgan" + 0.019*"aplikasi" + 0.019*"butuh"
Topic 6	0.041*"gabisa" + 0.041*"pas" + 0.028*"udh" + 0.025*"nya" + 0.025*"kali" + 0.023*"update" + 0.022*"atur" + 0.019*"masuk" + 0.017*"g" + 0.015*"gk"
Topic 7	0.155*"" + 0.098*"langgan" + 0.044*"ribet" + 0.032*"susah" + 0.025*"mini" + 0.023*"lagu2" + 0.022*"daftar" + 0.019*"potong" + 0.016*"ngk" + 0.013*"aktif"
Topic 8	0.204*"iklan" + 0.060*"nya" + 0.046*"lagu" + 0.029*"bintang" + 0.027*"bagus" + 0.027*"1" + 0.024*"tolong" + 0.023*"kasih" + 0.023*"iya" + 0.018*"2"
Topic 9	0.094*"mantap" + 0.091*"prem" + 0.038*"apl" + 0.030*"jarang" + 0.029*"isi" + 0.026*"bgus" + 0.020*"ngebug" + 0.016*"selera" + 0.016*"harga" + 0.015*"tapi"
Topic 10	0.099*"suara" + 0.028*"ko" + 0.027*"serba" + 0.026*"masalah" + 0.023*"hp" + 0.021*"pakai" + 0.016*"pulsa" + 0.015*"jernih" + 0.013*"nya" + 0.013*"ilangin"
Topic 11	0.098*"podcast" + 0.053*"baguss" + 0.048*"nikmat" + 0.034*"mood" + 0.032*"terkadang" + 0.031*"sihh" + 0.024*"i" + 0.023*"rem" + 0.019*"up" + 0.017*"rusak"
Topic 12	0.117*"lagu" + 0.091*"suka" + 0.042*"playlist" + 0.035*"spotify" + 0.034*"gk" + 0.029*"ny" + 0.026*"aplikasi" + 0.024*"play" + 0.021*"henti" + 0.019*"putar"
Topic 13	0.166*"lagu" + 0.102*"bagus" + 0.081*"aplikasi" + 0.080*"nya" + 0.047*"dengar" + 0.037*"musik" + 0.035*"suka" + 0.030*"cari" + 0.030*"spotify" + 0.022*"lengkap"
Topic 14	0.140*"lirik" + 0.068*"lagu" + 0.048*"tolong" + 0.046*"nya" + 0.046*"spotify" + 0.044*"lihat" + 0.035*"baik" + 0.026*"iya" + 0.019*"update" + 0.018*"muncul"
Topic 15	0.212*"musik" + 0.072*"dengerin" + 0.062*"dikit" + 0.048*"aplikasi" + 0.037*"premium" + 0.027*"dengar" + 0.020*"spotify" + 0.020*"enak" + 0.013*"denger" + 0.011*"biar"
Topic 16	0.156*"premium" + 0.111*"lagu" + 0.032*"spotify" + 0.028*"batas" + 0.023*"aplikasi" + 0.021*"putar" + 0.021*"pilih" + 0.019*"bayar" + 0.018*"iklan" + 0.016*"sih"

3.5 Discussion

The results of this study indicate that SVM outperforms NB in text classification, particularly in sentiment analysis of Spotify. These findings are consistent with research conducted by Leandro and Fianty (2025), which states that SVM provides better performance in terms of accuracy, precision, recall, and F1-score when compared to NB. The advantage of SVM lies in its ability to overcome class imbalance problems without requiring complex resampling methods. In addition, SVM is more sensitive to the complex characteristics of text data and the interdependent relationships between words. Meanwhile, NB, with its assumption of feature independence, experiences a decline in performance, particularly in detecting negative sentiment, as evidenced by its lower recall value. This shows that although NB is quite popular in text analysis, it has limitations in handling the complexity of word relationships in real text data [22].

The results of applying LDA to topic modeling produced 16 topics related to Spotify's serviceability dimension. Based on the analysis results, several key topics emerged from Spotify user reviews, including connectivity and accessibility, performance and user experience, premium services, application quality, content and playlists, application features, and sound and music quality. These findings highlight areas of significant concern to users, providing a basis for Spotify to improve and optimize the user experience. Therefore, several strategic steps are recommended for improvement, including upgrading server infrastructure, proactive server maintenance, and implementing AI-based chat support to provide quick and contextual solutions to problems frequently encountered by users.

However, although this study provides valuable insights, several limitations need to be considered. First, the data used is limited to user comments that can be accessed from certain platforms, which may not cover the entire picture of user experiences on other social media, such as X (Twitter), Instagram, or TikTok. Second, the NB model used is not optimal in handling feature dependencies, which limits its ability to detect more complex sentiments, especially negative ones. Thus, this study suggests that future research should expand the scope of data by collecting comments from more social media platforms and consider using other machine learning techniques, such as Neural Networks, to improve the accuracy of sentiment analysis, given the ever-evolving complexity of data.

4. CONCLUSION

The findings of this study demonstrate that the SVM algorithm surpasses NB in analyzing user sentiment related to Spotify's satisfaction. SVM achieved an accuracy rate of 87%, a precision of 87%, a recall of 81%, an F1-score of 84%, and an AUC score of 0.93, which validates SVM as a robust and efficient classification method.

Topic modeling using Latent Dirichlet Allocation (LDA) successfully identified various themes within the serviceability dimension, including connection and accessibility, application performance, premium services, content and playlist quality, application features, and sound and music quality. These findings provide valuable insights into the most frequently expressed concerns users have when evaluating Spotify's service.

As a follow-up, several recommendations are proposed to enhance user satisfaction, including server infrastructure upgrades, proactive server maintenance, the implementation of AI-powered chat support, personalized ad experience control, and improvements to the lyrics database. The implementation of these recommendations is expected to improve customer satisfaction, strengthen user retention, and enhance the overall user experience on the Spotify platform.

REFERENCES

[1] "Number of smartphone users in Indonesia from 2019 to 2029 (in millions). Statista. 2024," Statista.com. [Online]. Available: https://www.statista.com/forecasts/266729/smartphone-users-in-indonesia?srsltid=AfmBOopd3MuSzEK62T70cYdpG3LbwP8vawD6fiLG-_xhEYLYk_6wAsjM

[2] N. Adisty, "Mengulik perkembangan penggunaan smartphone di Indonesia," goodstats.id. [Online]. Available: https://goodstats.id/article/mengulik-perkembangan-penggunaan-smartphone-di-indonesia-sT2LA

[3] C. M. Annur, "Durasi penggunaan ponsel di Indonesia cenderung meningkat semenjak pandemi," databoks.katadata.co.id. [Online]. Available: https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/147c4723c1d145f/durasi-penggunaan-ponsel-di-indonesia-cenderung-meningkat-semenjak-pandemi

[4] A. S. Rahayu and A. Fauzi, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," vol. 4, pp. 349–354, 2022, doi: 10.30865/json.v4i2.5398.

[5] N. Naurah, "Bukan Spotify, Youtube Music Juarai Platform Musik yang Paling Sering Digunakan Masyarakat 2023," goodstats.id. [Online]. Available: https://goodstats.id/article/bukan-spotify-youtube-music-juarai-platform-musik-yang-paling-sering-digunakan-masyarakat-2023-o2Uyq

[6] D. Noviani, R. Pratiwi, S. Silvianadewi, M. Benny Alexandri, and M. Aulia Hakim, "Pengaruh Streaming Musik Terhadap Industri Musik di Indonesia," J. Bisnis Strateg., vol. 29, no. 1, pp. 14–25, 2020, doi: 10.14710/jbs.29.1.14-25.

[7] D. P. Sepnandito and Suharyati, "Pengaruh Fitur Spotify Social Dan PemasaranPersonalisasi Terhadap Loyalitas PelangganSpotify Dimediasi Dengan Kepuasan Pelanggan," J. Young Entrep., vol. 3, pp. 49–65, 2024, [Online]. Available: https://ejournal.upnvj.ac.id/index.php/jye

[8] H. A. Pradana, P. Pujiati, and N. Nina, "Hubungan Kualitas Pelayanan, Promosi dan Kepuasan Pelanggan terhadap Loyalitas Pelanggan Griya Sehat Kementerian Kesehatan Tahun 2021," J. Public Heal. Educ., vol. 3, no. 2, pp. 42–50, 2024, doi: 10.53801/jphe.v3i2.178.

[9] H. S. Le et al., "Predictive model for customer satisfaction analytics in E-commerce sector using machine learning and deep learning," Int. J. Inf. Manag. Data Insights, vol. 4, no. 2, p. 100295, 2024, doi: 10.1016/j.jjimei.2024.100295.

[10] K. S. Chong and N. Shah, "Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 12, pp. 90–94, 2022, doi: 10.14569/IJACSA.2022.0131213.

[11] M. A. Rahman and M. H. Seddiqui, "Comparison of Classical Machine Learning Approaches on Bangla Textual Emotion Analysis," pp. 2–7, 2019, [Online]. Available: http://arxiv.org/abs/1907.07826

[12] M. Guia, R. R. Silva, and J. Bernardino, "Comparison of Naive Bayes, support vector machine, decision trees and random forest on sentiment analysis," IC3K 2019 - Proc. 11th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag., vol. 1, no. Ic3k, pp. 525–531, 2019, doi: 10.5220/0008364105250531.

[13] A. Juwaini et al., "The role of customer e-trust, customer e-service quality and customer e-satisfaction on customer e-loyalty," Int. J. Data Netw. Sci., vol. 6, no. 2, pp. 477–486, 2022, doi: 10.5267/j.ijdns.2021.12.006.

[14] S. Asnaniyah, "Pengaruh E-Service Quality, E-Trust Dan E-Satisfaction Terhadap E-Loyalty Konsumen Muslim," J. Compr. Islam. Stud., vol. 1, no. 2, pp. 275–302, 2022, doi: 10.56436/jocis.v1i2.142.

[15] A. Farhan AlShammari, "Implementation of Keyword Extraction using Term Frequency-Inverse Document Frequency (TF-IDF) in Python," Int. J. Comput. Appl., vol. 185, no. 35, pp. 975–8887, 2023.

[16] N. Yusliani, S. A. Q. Aruda, M. D. Marieska, D. M. Saputra, and A. Abdiansah, "The effect of Chi-Square Feature Selection on Question Classification using Multinomial Naïve Bayes," Sinkron, vol. 7, no. 4, pp. 2430–2436, 2022, doi: 10.33395/sinkron.v7i4.11788.

[17] I. Darmawan, O. N. Pratiwi, F. R. Industri, and U. Telkom, "Analisis Sentimen Ulasan Produk Toko Online Rubylicious Untuk," e-Proceeding Eng., vol. 7, no. 2, pp. 7026–7034, 2020.

[18] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," J. Transform., vol. 18, no. 1, pp. 71–80, 2020, doi: 10.26623/transformatika.v18i1.2317.

[19] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," Teknika, vol. 10, no. 1, pp. 18–26, 2021, doi:

10.34148/teknika.v10i1.311.
[20] S. Karmila and V. I. Ardianti, “Metode Latent Dirichlet Allocation Untuk Menentukan Topik Teks Suatu Berita,” J. Inform. dan Komputasi Media Bahasan, Anal. dan Apl., vol. 16, no. 01, pp. 36–44, 2022, doi: 10.56956/jiki.v16i01.100.
[21] T. Titiana and D. H. Bangkalang, “Analisis Dan Penerapan Topic Modeling Pada Judul Tugas Akhir Mahasiswa Menggunakan Metode Latent Dirichlet Allocation (Lda),” JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 8, no. 4, pp. 1275–1287, 2023, doi: 10.29100/jipi.v8i4.4254.
[22] J. O. Leandro and M. I. Fianty, “Evaluation of Sentiment Analysis Methods for Social Media Applications: A Comparison of Support Vector Machines and Naïve Bayes,” Int. J. Informatics Vis., vol. 9, no. 2, pp. 796–807, 2025, doi: 10.62527/joiv.9.2.2905.
[23] F. S. Nahm, “Receiver operating characteristic curve : overview and practical use for clinicians,” 2022.
[24] S. Syed and M. Spruit, “Exploring Symmetrical and Asymmetrical Dirichlet Priors for Latent Dirichlet Allocation,” Int. J. Semant. Comput., vol. 12, no. 3, pp. 399–423, 2018, doi: 10.1142/S1793351X18400184.
[25] M. Röder, A. Both, and A. Hinneburg, “Exploring the space of topic coherence measures,” WSDM 2015 - Proc. 8th ACM Int. Conf. Web Search Data Min., pp. 399–408, 2015, doi: 10.1145/2684822.2685324.
[26] G. Satria, P. Ramadhan, and S. Nugroho, “Optimasi Data Preprocessing dan Hyperparameter Tuning pada Klasifikasi Penyakit Daun Apel menggunakan DenseNet169,” vol. 6, no. 3, pp. 1352–1362, 2024, doi: 10.47065/bits.v6i3.6134.
[27] “Ideal Dataset Splitting Ratios In Machine Learning Algorithms :,” pp. 496–504, 2021.
[28] H. Bichri, A. Chergui, and M. Hain, “Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets,” Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 2, pp. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.
[29] Y. Nuri and E. Senyurek, “Research Abstracts Similarity Implementation By Using TF-IDF Algorithm,” vol. 27, no. 1, pp. 4–10, 2025, doi: 10.9790/0661-2701040410.
[30] X. Cheng, “A Comprehensive Study of Feature Selection Techniques in Machine Learning Models,” Insights Comput. Signals Syst., vol. 1, no. 1, pp. 65–78, 2024, doi: 10.70088/xpf2b276.
[31] S. F. Sandita, I. T. Utami, and M. Y. Rochayani, “Sentiment Classification of Livin by Mandiri App Reviews with Support Vector Machine Based on Modified Particle Swarm Optimization,” Int. J. Math. Comput. Res., vol. 13, no. 05, pp. 5260–5266, 2025, doi: 10.47191/ijmcr/v13i5.18.
[32] N. Fajriyah, N. T. Lapatta, D. W. Nugraha, and R. Laila, “Implementasi Svm Dan Smote Pada Analisis Sentimen Media Sosial X Terhadap Pelantikan Agus Harimurti Yudhoyono,” JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform., vol. 10, no. 2, pp. 1359–1370, 2025, doi: 10.29100/jipi.v10i2.6246.
[33] R. D. Fitriani, H. Yasin, and T. Tarno, “Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal),” J. Gaussian, vol. 10, no. 1, pp. 11–20, 2021, doi: 10.14710/j.gauss.v10i1.30243.

BIBLIOGRAPHY OF AUTHORS

Mutawakkil Samjas is an industrial engineering student at Hasanuddin University, Makassar, Indonesia. His research focuses on data analysis.

Armin Darmawan, Ph.D. is currently an Associate Professor in the Department of Industrial Engineering, Hasanuddin University, Makassar, Indonesia. His research focuses on Quality Management, Statistical Quality Control, Data Analysis, Risk Management, and Decision Analysis.