Journal of Information Systems and Informatics
Vol. 7, No. 1, March 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i1.1041 Published By DRPM-UBD

Harnessing SVM for Sentiment Analysis: Insights from Gojek's Instagram Engagement

Muhammad Juan Savero[1], Ali Ibrahim[2]*, Yadi Utama[3], Endang Lestari[4]

[1,2,3,4] Information System, Faculty of Computer Science, Sriwijaya University, Indonesia
[2] Master of Computer Science, Faculty of Computer Science, Sriwijaya University, Indonesia
Email: [1]juansavero27@gmail.com, [2*]aliibrahim@unsri.ac.id, [3]yadiutama@unsri.ac.id, [4]endanglestari@unsri.ac.id

Abstract

The development of digital technology has changed the transportation industry, including online services such as Gojek. Understanding customer sentiment is key in improving user experience and designing more effective business strategies. This research analyzes Gojek user sentiment on Instagram using Support Vector Machine (SVM). Data is obtained through web scraping, then processed through text cleaning, tokenization, common word removal, and stemming. Features were extracted using Term Frequency-Inverse Document Frequency (TF-IDF) before being classified with SVM. The results showed that the SVM model achieved 70.82% accuracy in classifying user sentiment. Most positive comments highlight the convenience and efficiency of the service, while negative comments are more related to high tariffs, application constraints, and less responsive customer service. These findings provide insights for Gojek to improve marketing strategies, optimize customer service, and adjust fare policies based on user feedback. In addition, this analysis can help in predicting real-time customer satisfaction trends through sentiment monitoring on social media. As a development step, this research recommends further exploration with deep learning and Aspect-Based Sentiment Analysis (ABSA) to improve accuracy and understand the service aspects that have the most influence on customer satisfaction.

Keywords: Sentiment Analysis, Support Vector Machine, Gojek, Instagram, Marketing Strategy, Customer Satisfaction Prediction

1. INTRODUCTION

The development of digital technology has brought major changes in various sectors, including the transportation industry. App-based transportation services such as Gojek have become the main solution for urban communities in meeting their mobility needs [1]. By simply using an application on a smartphone, users can access a variety of services, ranging from transportation to delivery of goods and food [2]. This practicality makes Gojek one of the most popular online transportation platforms in Indonesia. However, as the number of users increases, the challenge of understanding customer satisfaction and perceptions of this service is an important aspect that needs to be analyzed so that companies can continue to improve their service quality [3].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

One effective way to evaluate customer satisfaction is through data-driven sentiment analysis from social media, such as Instagram. Users' comments on these platforms often reflect their real-life experiences with the services used [4]. However, sentiment analysis has its own challenges, such as processing informal language, slang, and sarcasm, which can make it difficult for automated systems to accurately interpret customer opinions [5]. Previous research generally uses Naïve Bayes, K-Nearest Neighbors (KNN), or Random Forest methods for sentiment analysis, but these methods have limitations in handling complex text data [6]. Therefore, this research uses Support Vector Machine (SVM), which is known to perform better in high-dimensional text classification and produces more optimal accuracy than other methods [7].

This research aims to analyze Gojek user sentiment on Instagram with the SVM algorithm, classify customer opinions into positive and negative categories, and evaluate the performance of the model used. In addition, this research also identifies the main themes in user comments, such as tariffs, service quality, and technical application constraints, which can be a key indicator in understanding customer satisfaction. The results of this study are expected to help Gojek in adjusting its data-driven business and marketing strategies, so that the company can be more responsive to customer feedback and improve its service quality [8].

Sentiment analysis has been applied in various industries to improve business strategies and customer experience. In the banking sector, sentiment analysis is used to measure customer satisfaction with digital services and customer service . In the e-commerce industry, this method has helped companies predict customer shopping trends and adjust marketing strategies in real-time [9]. Based on these benefits, the application of sentiment analysis in online transportation services such as Gojek can help in identifying the factors that have the most influence on user experience and optimizing customer service [10]. By understanding customer sentiment trends more accurately, companies can design data-driven strategies to increase user loyalty and business competitiveness.

2. METHODS

The research method applied includes a series of systematic steps that serve as a guide or work plan to facilitate the achievement of expected results. All stages of the research were carried out using Google Collaboratory, a cloud-based platform that allows the execution of Python code directly in the browser without requiring software installation on the computer. An illustration of the stages of this research is presented in Figure 1.

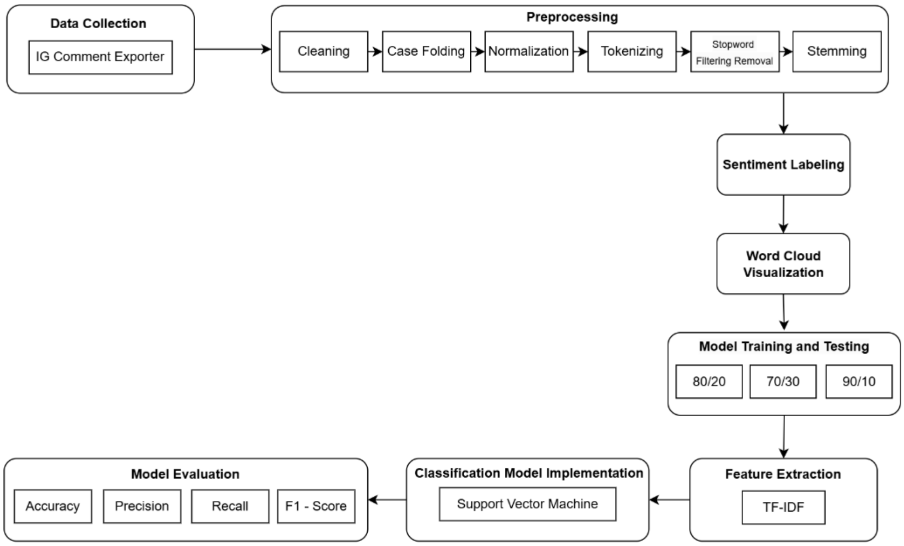


Figure 1. Research Stages

2.1. Data Collection

Online review data collection is done through web scraping using the Google Chrome extension IG Comment Exporter. This technique allows automatic data extraction from Instagram without relying on APIs that have limited access. This makes it easy to collect different types of data, such as user comments, engagement metrics, and trends, which are useful for research and analysis purposes.

2.2. Preprocessing

Preprocessing is the initial stage in data processing to organize data in a structured manner and reduce noise, thus speeding up computation. In this research, text preprocessing is done to analyze sentiment more accurately. This process includes cleaning, case folding, normalization, tokenization, stopword removal, and stemming.

2.3. Sentiment Labeling

Data labeling is done manually using Microsoft Excel, starting swith collecting comments from Gojek's official account on Instagram. After that, the data was processed through text cleaning, tokenization, and stemming stages to make it more structured [11]. Next, researchers manually classified each comment into positive or negative sentiment categories based on an understanding of the context and language usage by users. To ensure accuracy and consistency in labeling, the

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

classification results have been verified by a lecturer who is an expert in the field of language and text analysis [12]. With this validation, it is expected that the analysis results can reflect the right sentiment in accordance with the social and cultural context of Gojek users, thus increasing the credibility and quality of the findings in this study.

2.4. Word Cloud Visualization

The results of this sentiment analysis will be visualized with a word cloud to display the words that appear most often in positive and negative sentiments. This visualization makes it easy to identify the main topics discussed by Gojek users and will be further analysed to provide recommendations to developers.

2.5. Model Training and Testing

In the analysis of Gojek user reviews, the dataset is divided into three scenarios, namely 70:30, 80:20, and 90:10, using the sklearn library. The stemming result column is used as the feature, while the label column as the target. After the division, the dataset size was checked to ensure the appropriate proportion. This comparison was done to see the impact on model performance and determine the best ratio for analysis.

2.6. Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a text analysis method that assesses the importance of a word in a document compared to the entire corpus. TF is calculated based on how often a word appears in a document, while IDF measures how rarely the word is found in all documents [13]. The calculation formula is as shown in Equation 1 and 2.

$$\mathrm{TF(t,d)}=\frac{\text{Number of occurrences of term t in document d}}{\text{Total number of terms in the document d}} \quad (1)$$

$$\mathrm{IDF(t)}=\log\left(\frac{N}{\mathrm{DF(t)}}\right) \quad (2)$$

where N is the total number of documents, and DF(t) denotes the number of documents containing word t. TF-IDF is widely used in text classification, document similarity measurement, and text clustering. In clustering, this method is often applied in algorithms such as K-Means to group documents based on content similarity [13]. In addition, TF-IDF is also used together with the concept of cosine similarity to analyze the topic similarity between documents. In text classification, TF-IDF is proven to provide competitive results compared to other feature extraction methods [9]. However, this technique has limitations in

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

capturing the semantic meaning between words, so it is often combined with machine learning-based methods such as word embeddings or deep learning [14].

2.7. Classification Model Implementation

After the dataset preparation process is complete, this research uses Support Vector Machine (SVM) as a classification algorithm for sentiment analysis. SVM was chosen due to its ability to handle high-dimensional data and is effective in text classification [9]. This algorithm works by finding the optimal hyperplane that separates sentiment categories and maximizes the margin between data. Based on the results of this study, the SVM model achieved an average accuracy of 70% in classifying the sentiment of user comments. Although this figure is not optimal, SVM is still chosen because of its reliability in handling large text data with complex patterns. Previous studies have shown that SVM performance is greatly influenced by the feature selection and preprocessing techniques used [15]. When compared to other algorithms such as Naïve Bayes and Random Forest, SVM has an advantage in handling unstructured and high-dimensional text data [10]. However, this study found that SVM has an accuracy of about 70%, while another study showed that:

1) Naïve Bayes has an accuracy of around 65%-75%, although it has better processing speed.
2) Random Forest generally has an accuracy between 72%-80%, but requires longer computation time [10].

One of the main challenges in applying SVM to social media sentiment analysis is the difficulty in recognizing sarcasm, slang, and ambiguous comments that often appear in online conversations. SVM-based models sometimes lack accuracy in capturing the implicit meaning of text, especially if they are not supported by a training dataset that includes such language variations [16]. Therefore, to improve model performance, this study recommends hyperparameter optimization and the application of feature engineering to improve classification accuracy. Some studies have also shown that the combination of SVM with advanced feature extraction techniques, such as Enhanced Vector Space Model (EVSM) or Latent Semantic Analysis (LSA), can improve accuracy up to 92.78% in sentiment analysis [17]. Mathematically, the hyperplane in SVM is used to separate sentiment categories with Equation 3.

$$w0 + w1\ x1 + w2\ x2 = 0 \tag{3}$$

Description:

$w0$	:	bias that determines the position of the hyperplane from the origin.
$w1$ and $w2$	:	weights for features x1 and x2

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

$x1$ and $x2$: input variable specifies the position of a point in two dimensionalspace

Data above the hyperplane ($w0+ w1\ x1 + w2\ x2 > 0$) is categorized as positive sentiment, while data below it ($w0 + w1\ x1 + w2\ x2 < 0$) is categorized as negative sentiment [10]. In this study, although SVM accuracy only reached 70%, this model was still chosen due to its stability in handling complex datasets [18]. To improve performance, future research can consider further use of hyperparameter tuning as well as exploration of deep learning-based models to improve context understanding in sentiment analysis.

2.8. Model Evaluation

The last stage in sentiment analysis using Support Vector Machine (SVM) is model evaluation, which aims to measure the extent to which the model can classify sentiment accurately after going through the training and testing process [19]. This evaluation is done using several metrics, such as accuracy, precision, recall, and F1 score, which are obtained from Confusion Matrix. Confusion Matrix itself is used to see the number of correct and incorrect predictions in determining positive and negative sentiments [20], [21]. In Google Collaboratory, the evaluation process is carried out by utilizing the sklearn.metrics library, which provides various functions to calculate evaluation metrics based on the model prediction results against test data. The evaluation results are then further analyzed to determine the extent to which the model is able to classify sentiment properly. If the model's performance is still not optimal, several improvement steps can be taken, such as adjusting SVM parameters, cleaning the training dataset, or using cross validation techniques to make the model more accurate [22][23]. With this evaluation stage, the model that has been developed is expected to produce sentiment analysis that is more accurate, reliable, and in accordance with research needs.

3. RESULTS AND DISCUSSION

This section presents the research results obtained based on the methodology described in the previous section, starting from the results of data mining to the model evaluation process.

3.1. Data Collection

The data scraping process is done using the extension feature of the Goggle Chrome application, namely IG Comment Exporter. The stages of the Data Scraping process are presented in Figure 2.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Figure 2. Data Collection Process

From the results of this Data Scraping, data consisting of 1,645 comments from Gojek Indonesia's Instagram account were collected, which were then used as raw data for further analysis processes in this study. The following is an example of the results of Data Scraping.

Table 1. Sample Data Collection (Indonesia)

Comments
pakegojekpalinghemat.
Tolong Cek dm Saya, Refund gomart saya Belum masuk Rek fitur bantuan di apk Gojek tidak membantu sama Sekali Jawabannya semua template.
lebih hemat naik Gocar.
ini kok lama sekali ya ngirim gosend nya? driver nya Pada kemana.
Jancok @gofood.
asyik Cobain gocar.

3.2. Preprocessing

The data preprocessing stage was carried out to prepare the comment data before model implementation. This stage involved several essential processes, including cleaning, case folding, normalization, tokenizing, stopword removal (filtering), and stemming. The preprocessing was conducted using collaborative tools such as Google Sheets for efficiency and accuracy. The results of this preprocessing stage are presented in Table 2, which illustrates the transformation of raw text into structured and cleaned data ready for sentiment analysis.

Table 2. Preprocessing (Indonesia)

Comments	
Original	ini kok lama sekali ya ngirim gosend nya? driver nya Pada kemana.
cleaning	ini kok lama sekali ya ngirim gosend nya driver nya Pada kemana
case folding	ini kok lama sekali iya mengirim gosend nya driver nya pada kemana

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

	Comments
normalization	ini kok lama sekali iya mengirim gosend nya driver nya pada kemana
tokenizing	['ini', 'kok', 'lama', 'sekali', 'iya', 'mengirim', 'gosend', 'nya', 'driver', 'nya', 'pada', 'kemana']
stopword or filtering removal	['iya', 'mengirim', 'gosend', 'nya', 'driver', 'nya', 'kemana']
stemming	['iya', 'kirim', 'gosend', 'nya', 'driver', 'nya', 'mana']

3.3. Data Labeling

In the classification process, researchers manually assessed each comment based on context and language usage to determine positive or negative sentiment. To maintain accuracy and consistency, the labeling results were verified by one of Sriwijaya University's Faculty of Computer Science lecturers who is an expert in language and text analysis. Validation is done by reviewing a random sample of data, comparing the labeling results with linguistic guidelines, and identifying inconsistencies. If discrepancies were found, discussions and revisions were made to improve data reliability. After validation, the final data was used in sentiment analysis. Previously, 1,648 data were preprocessed and manually classified into 778 positive sentiments and 869 negative sentiments using Microsoft Excel.

One of the main challenges in analyzing social media data is handling ambiguous comments. Comments that are sarcastic, ironic or have double meanings are often difficult to classify accurately, especially when there is no clear additional context. In addition, the use of informal language, slang or code-switching can make it difficult to interpret the true sentiment. To address this, researchers conducted in-depth discussions with linguists and developed more detailed classification guidelines to keep the labeling consistent.

Table 3. Sample Labelling Data (Indonesia)

Comment	Labelling
pakegojekpalinghemat.	positive
tolong Cek dm Saya, Refund gomart saya Belum masuk Rek fitur bantuan di apk	negative
lebih hemat naik Gocar.	positive
ini kok lama sekali ya ngirim gosend nya? driver nya Pada kemana.	negative
Jancok @gofood	negative
asyik Cobain gocar.	positive

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

3.4. Word Cloud Visualization

This research aims to provide a deeper understanding of the responses and perceptions of Gojek users in Indonesia. To achieve this goal, the analysis results are presented in two separate sections, one section for reviews with positive sentiment and another for reviews containing negative sentiment. Figure 3 displays a wordcloud depicting positive responses from Gojek users. Words such as “gojek”, “driver”, and “good” appear dominant, indicating that users highly appreciate the driver's service as well as the overall service quality. In addition, the occurrence of the words “really”, “practical”, “fast”, “save”, and “road” reflect the level of customer satisfaction with the aspects of convenience, efficiency, and economic benefits provided. Other words such as “gocar”, “gosend”, “gofood”, and “gomart” also appear frequently, indicating that Gojek's various features and services have received positive responses from its users.

Figure 3. Positive Reviews Word Cloud

Figure 4. Negative Review Word Cloud

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 4 presents a wordcloud illustrating the various complaints of Gojek users in their negative reviews. Words such as "gojek", "driver", and "admin" dominate, indicating that most problems are related to driver services and customer support. In addition, the occurrence of the words "application", "login", and "balance" indicate problems in using the system, especially related to the login and transaction process. Words such as "please", "report", and "cancel" also appear frequently, indicating that users face difficulties in accessing help and experience problems when booking or canceling services.

3.5. Model Training and Testing

The processed dataset was divided into three split scenarios, namely 70:30, 80:20, and 90:10, to analyze the effect of data proportion on model performance. This division is done so that the model can be optimally trained as well as tested in a more objective way. In each scenario, most of the data is used for the training process, while the rest is used for testing. In the 70:30 scenario, the model has more test data to measure performance more broadly. Meanwhile, the 80:20 scenario provides a balance between training and testing data, while the 90:10 scenario allows the model to get more training data, but with less test data. The figure below shows the data distribution after the split. In each scenario, the amount of training data is always more than the testing data. In addition, reviews with negative sentiment remain more dominant than positive sentiment in all scenarios. The visualization of the data division results can be seen in Figure 5.

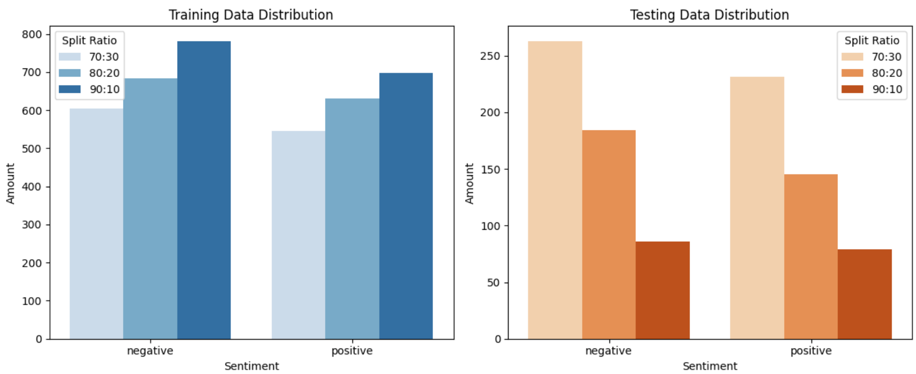


Figure 5. Visualization Of Training and Testing Models

3.6. Feature Extraction

The TF-IDF method is applied to assign weights to each word in the text, thus enabling more effective identification of important terms. This weighting technique determines words that frequently appear in the reviews but are rarely found in the whole corpus, thus helping in understanding the user's perspective

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

more deeply. Table 4 displays the TF-IDF value for each term contained in the Gojek user comment dataset in Indonesia.

Table 4. TF-IDF

Feature	TI-IDF	Document
bayar	0.313793	0
applicator	0.276256	0
lunasin	0.276256	0
november	0.276256	0
pinjam	0.276256	0
hemat	0.723067	1
pakai	0.584778	1
gojek	0.367708	1
bantu	0.409656	2
rek	0.373353	2

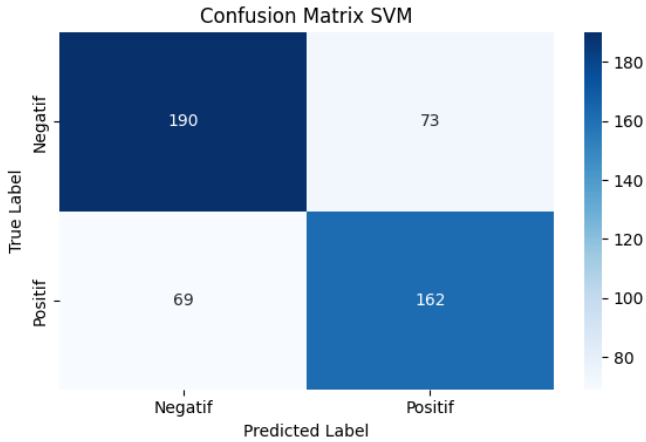
3.7. Classification Model Implementation

The SVM algorithm is implemented using the LinearSVC class from the scikit-learn library, which is designed for Support Vector Classification in the case of linear splitting. Linear kernels were chosen because text data that has undergone TF-IDF transformation is generally linearly separable. The model is trained using the processed training data, with the aim of categorizing the reviews into two groups, namely positive and negative. Further explanation of the SVM algorithm implementation will be explained in the next section.

3.8. Model Evaluation

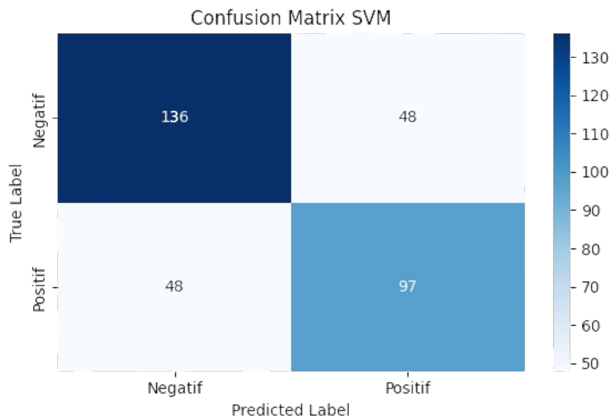
Model performance will be analyzed using a confusion matrix. This evaluation is done to measure the effectiveness of each algorithm in classifying sentiments contained in Gojek user comments in Indonesia, both positive and negative. Figure 6 the sentiment classification results performed by the SVM model in the 70:30 ratio scenario. The model correctly identified 190 negative reviews and 162 positive reviews. However, there were 73 negative reviews that were incorrectly categorized as positive, and 69 positive reviews that were classified as negative. This result indicates that the model performs well, but there is still an opportunity to improve accuracy, especially in reducing misclassification.

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882



Gambar 6. Data Separation Visualization of 70:30 Ratio

Figure 7 is the prediction results of the Support Vector Machine (SVM) model with an 80:20 comparison scenario in classifying the sentiment of Gojek user reviews. Based on test data, the SVM model is able to classify 136 negative reviews and 97 positive reviews correctly. However, there are still some prediction errors, where 40 negative reviews are categorized as positive, while 48 positive reviews are categorized as negative. These results show that the SVM model has a fairly good performance in distinguishing positive and negative sentiments in Gojek user reviews.

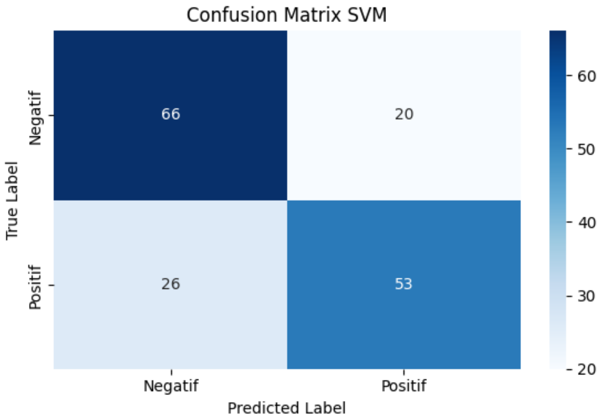


Gambar 7. Data Separation Visualization of 80:20 Ratio

Figure 8 displays the sentiment classification results of Gojek user reviews using the Support Vector Machine (SVM) model with a 90:10 comparison scenario. From the test results, the model was able to correctly identify 66 negative reviews and 53 positive reviews. However, there are still prediction errors, where 20

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

negative reviews are classified as positive, and 26 positive reviews are incorrectly categorized as negative. This finding shows that the SVM model performs quite well in distinguishing sentiment, although there are still errors in classification.



Gambar 8. Data Separation Visualization of 90:10 Ratio

To assess the performance of the SVM model in classifying the sentiment of Gojek user reviews, refer to Table 5. This table presents the main evaluation metrics, namely accuracy, precision, recall, and F1 score, with three data sharing scenarios: 70:30, 80:20, and 90:10. In the 70:30 data split, the SVM model achieved 71.26% accuracy, with 71.15% precision, 71.19% recall, and 71.16% F1 score. Meanwhile, in the 80:20 scenario, the accuracy obtained was 70.82%, with precision, recall, and F1 score values of 70.40% each. In the 90:10 scenario, the model showed the best performance with 72.12% accuracy, 72.17% precision, 71.92% recall, and an F1 score of 71.95%. These results indicate that the SVM model is able to classify sentiment quite well, with increased accuracy in the 90:10 data sharing scenario.

Table 5. Model Performance

Model	Comparison	Accuracy	Precision	Recall	F-1 Score
SVM	70 : 30	71.26%	71.15%	71.19%	71.16%
SVM	80 : 20	70.82%	70.40%	70.40%	70.40%
SVM	90 : 10	72.12%	72.17%	71.92%	71.95%

Figure 9 shows the performance comparison of the Support Vector Machine (SVM) model on three data sharing scenarios, namely 70:30, 80:20, and 90:10, based on accuracy, precision, recall, and F1-score metrics. The results show that the model with the 90:10 scenario has the best performance with an accuracy of about 72.12%, followed by the 70:30 scenario with an accuracy of 71.26%, and the 80:20 scenario with an accuracy of 70.82%. The precision, recall, and F1-score values also follow a similar pattern, with the 90:10 scenario showing higher

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

performance than the other scenarios. This graph indicates that the larger the proportion of training data, the model tends to have better and more stable performance in classifying the sentiment of Gojek user reviews.

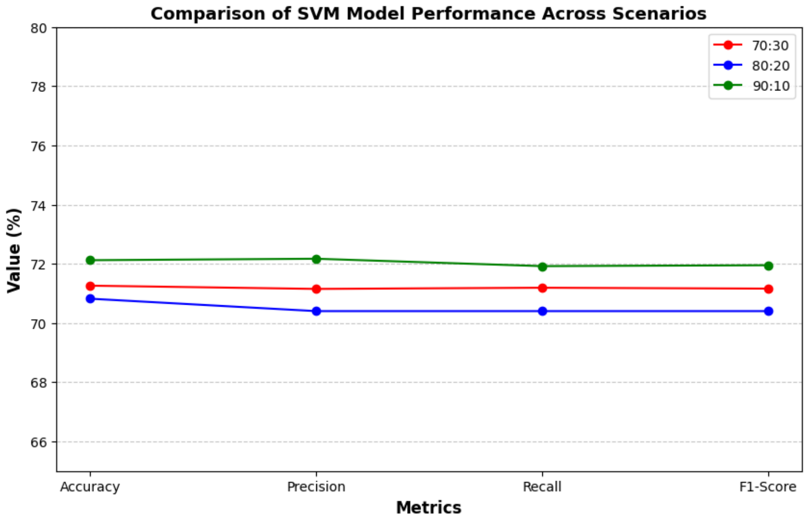


Figure 9. Model Performance Graph

3.9. Discussion

The findings of this study highlight several critical themes influencing user sentiment, pointing to deeper systemic issues and opportunities within ride-hailing platforms such as Gojek. While the sentiment classification results confirmed the presence of both positive and negative perceptions, a more meaningful interpretation emerges when we synthesize these sentiments with known service experience factors and user expectations.

One of the most notable insights is the influence of fare transparency and pricing strategy on user trust and satisfaction. Although Gojek offers various promotions and discounts, which have a positive effect on user sentiment, the unpredictability of pricing and surge fares is a consistent source of dissatisfaction. This aligns with prior research suggesting that price instability can lead to user frustration and switching behavior [24]. The discrepancy between expected and actual costs creates a psychological barrier, where users feel unfairly treated even if they continue to use the service out of convenience or habit. This points to a need for more transparent fare estimation features and real-time price justifications within the app.

Another key factor is application performance, which directly affects usability and user experience. Comments expressing frustration with app crashes, delays, or

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

payment failures suggest that technical reliability remains a core concern. This is particularly important in the context of digital services, where app interface and backend performance are the only points of contact between the user and service provider. System errors do not just interrupt transactions—they erode confidence and diminish perceived service value. This echoes prior findings emphasizing that consistent performance is essential for user retention in app-based services [25]. It also underscores the strategic value of investing in mobile UX and backend stability.

Customer service interactions further influence sentiment, often amplifying initial user frustration. Users expect fast, personalized, and effective responses when issues arise. The prevalence of negative sentiment associated with unresolved complaints or generic replies signals a disconnect between user expectations and actual service delivery. As noted in earlier studies, customer support is not merely a reactive feature it is a brand touchpoint that can either recover trust or drive users away [25]. Integrating AI-driven support systems with human escalation paths could improve efficiency while maintaining empathy in response handling.

From a machine learning perspective, the study demonstrates that dataset balance plays a pivotal role in model accuracy and fairness. The relatively strong performance of the SVM model (especially in the 90:10 data split scenario) suggests that increasing the proportion of training data helps the model generalize better, a finding consistent with common practices in text classification. However, performance differences across the three split ratios also reveal the trade-off between model generalization and validation reliability. The model's ability to handle Gojek-specific vocabulary such as gofood, gosend, gopay further confirms the value of domain-specific preprocessing and TF-IDF feature extraction.

On a broader level, the study confirms that social media sentiment reflects not just customer experience, but broader perceptions of brand reliability, value delivery, and operational integrity. These sentiments are influenced not only by direct interactions, but also by public narratives, peer experiences, and historical service performance. Therefore, sentiment analysis should not be treated as a static evaluation, but as a continuous listening mechanism to capture evolving user expectations and guide service improvement.

The analysis reinforces the need for greater fare transparency, robust application performance, and responsive customer support. At the same time, it showcases the power of integrating natural language processing and supervised learning models to extract actionable insights from unstructured social data. Future work should explore incorporating context-aware language models to better interpret ambiguous expressions, sarcasm, or slang, which are prevalent in Indonesian social media. Moreover, ongoing sentiment tracking could serve as a real-time feedback system for continuous service optimization.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

4. CONCLUSIONS

This study demonstrates that the Support Vector Machine (SVM) algorithm is a viable approach for analyzing Gojek customer sentiment on social media, achieving an accuracy of 70.82% in classifying user comments into positive and negative categories. The analysis reveals that positive sentiments are primarily associated with the convenience of ordering, the efficiency of services, and driver professionalism. In contrast, negative sentiments tend to arise from unpredictable pricing, technical issues within the application, and delays or inadequacies in customer service response. To enhance customer satisfaction and loyalty, Gojek can adopt several targeted strategies. First, improving the responsiveness and personalization of customer support, possibly through the integration of AI-powered chatbots for handling initial inquiries, can significantly reduce user frustration. Second, ongoing investment in application stability and system performance is critical to minimizing service disruptions and enhancing user trust. Third, addressing concerns about fare transparency—particularly by offering clearer justifications for surge pricing and implementing loyalty incentives or discount schemes—can help users feel more fairly treated and valued.

For future development, this study suggests exploring more advanced models, such as Long Short-Term Memory (LSTM) or Bidirectional Encoder Representations from Transformers (BERT), which may provide greater contextual understanding and improved classification accuracy. Additionally, implementing Aspect-Based Sentiment Analysis (ABSA) would enable more granular insights by identifying which specific service components (e.g., payment, delivery time, app interface) most strongly influence customer sentiment. By combining strategic service improvements with advanced sentiment analysis, Gojek can not only respond to current user concerns but also proactively adapt to evolving customer expectations—ultimately fostering stronger engagement, trust, and competitive advantage in Indonesia's ride-hailing industry.

ACKNOWLEDGEMENT

Thank you to the Faculty of Computer Science for providing facilities for writing articles, especially the Information Systems Department of the Faculty of Computer Science, Sriwijaya University.

REFERENCES

[1] M. N. Margaretha dan D. H. Wibowo, "Hubungan antara kepuasan hidup dengan Fear of Missing Out (FoMO) pada remaja pengguna media sosial," *Psikoislamika: J. Psikol. dan Psikol. Islam*, vol. 20, no. 2, pp. 2655–5034, 2023, doi: 10.18860/psi.v20i2.23219.s.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[2] G. Y. Pratama dan A. Suradi, "Perlindungan hukum terhadap data pribadi pengguna jasa transportasi online dari tindakan penyalahgunaan pihak penyedia jasa berdasarkan Undang-Undang Nomor 8 Tahun 1999 tentang Perlindungan Konsumen," *Diponegoro Law J.*, vol. 5, no. 3, 2016.

[3] P. Yuniar dan Kismiantini, "Analisis sentimen ulasan pada Gojek menggunakan metode Naive Bayes," *Statistika*, vol. 23, no. 2, pp. 164–175, Dec. 2023, doi: 10.29313/statistika.v23i2.2353.

[4] Fathrizqy, P. Mahardika, A. Larasati, dan A. Muid, "The sentiment analysis of online customer review on food and beverages delivery services in the GOJEK application using K-Nearest Neighbors," in *Proc. 12th Int. Conf. Ind. Eng. Oper. Manag. (IEOM)*, Istanbul, Turkey, Mar. 2022, pp. 167–176.

[5] P. Heydarian, A. Bifet, dan S. Corbet, "Understanding market sentiment analysis: A survey," *J. Econ. Surv.*, vol. 38, no. 1, pp. 1–25, 2024, doi: 10.1111/joes.12645.

[6] B. F. Wiguna et al., "Indonesian Scientific Index (SINTA) journal-level of S3," *Piksel*, vol. 11, no. 2, p. 401, 2023, doi: 10.33558/piksel.

[7] J. J. Sihombing dan R. I. Sitorus, "Implementasi algoritma Support Vector Machine untuk analisis sentimen aplikasi OLX di Playstore," *J. Inform. Data Sci. (J-IDS)*, vol. 1, no. 2, 2022.

[8] H. Andreansyah, "Klasifikasi sentimen positif dan negatif pada ulasan aplikasi Gojek menggunakan metode Support Vector Machine (SVM)," *JURASIK*, vol. 9, no. 1, pp. 329–336, 2024.

[9] R. J. Bhuiyan et al., "Sentiment analysis of customer feedback in the banking sector: A comparative study of machine learning models," *Am. J. Eng. Technol.*, vol. 6, no. 10, pp. 54–66, Oct. 2024, doi: 10.37547/tajet/Volume06Issue10-07.

[10] B. A. Ardhani, N. Chamidah, dan T. Saifudin, "Sentiment analysis towards Kartu Prakerja using text mining with Support Vector Machine and Radial Basis Function Kernel," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 2, pp. 119–128, Oct. 2021, doi: 10.20473/jisebi.7.2.119-128.

[11] Y. Akbar dan T. Sugiharto, "Analisis sentimen pengguna Twitter di Indonesia terhadap ChatGPT menggunakan algoritma C4.5 dan Naïve Bayes," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 115–122, 2023, doi: 10.55338/saintek.v4i3.1368.

[12] I. Amal, "Perbandingan pelabelan otomatis dan manual untuk analisis sentimen terhadap kenaikan harga BBM Pertamina pada Twitter menggunakan algoritma Support Vector Machine," *J. Inform.*, vol. 8, no. 3, pp. 123–130, 2023.

[13] A. Widianto dan E. Pebriyanto, "Document similarity using Term Frequency-Inverse Document Frequency representation and cosine similarity," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 4, no. 2, pp. 149–153, 2024.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

[14] R. E. Putra, F. Pramudita, dan D. R. Anggraini, "Sentiment analysis terhadap layanan transportasi online menggunakan metode Naïve Bayes dan Support Vector Machine," *J. Teknol. dan Sist. Komput.*, vol. 11, no. 2, pp. 189–195, 2023, doi: 10.14710/jtsiskom.2023.189.

[15] A. J. N. Kisma, P. Arsi, dan P. Subarkah, "Sentiment analysis regarding candidate presidential 2024 using Support Vector Machine Backpropagation based," *JTAM (J. Teor. Apl. Mat.)*, vol. 8, no. 1, p. 96, Jan. 2024, doi: 10.31764/jtam.v8i1.17294.

[16] V. B. Lestari dan A. Dini, "Support Vector Machine for sentiment analysis of PT. Paragon Technology and Innovation (Case study of brand Make Over and Emina product users on Female Daily Page-Beauty Review)," in *Proc. Conf.*, 2024.

[17] K. S. Kumar et al., "Sentiment analysis of short texts using SVMs and VSMs-based multiclass semantic classification," *Appl. Artif. Intell.*, vol. 38, no. 1, 2024, doi: 10.1080/08839514.2024.2321555.

[18] B. Sugara dan A. Subekti, "Penerapan Support Vector Machine (SVM) pada small dataset untuk deteksi dini gangguan autisme," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 177–182, Sep. 2019, doi: 10.33480/pilar.v15i2.649.

[19] A. M. Yolanda dan R. T. Mulya, "Implementasi metode Support Vector Machine untuk analisis sentimen pada ulasan aplikasi Sayurbox di Google Play Store," *VARIANSI: J. Stat. dan Apl. dalam Pembelajaran dan Penelitian*, vol. 6, no. 2, pp. 76–83, 2024, doi: 10.35580/variansiunm258.

[20] B. Gunawan, H. Sasty, E. Esyudha, dan P. #3, "Sistem analisis sentimen pada ulasan produk menggunakan metode Naive Bayes," *JEPIN (J. Edukasi dan Penelit. Informatika)*, vol. 4, no. 2, pp. 17–29, 2018.

[21] I. Juventius, T. Gurning, P. P. Adikara, dan R. S. Perdana, "Analisis sentimen dokumen Twitter menggunakan metode Naïve Bayes dengan seleksi fitur GU Metric," *J. Ilm. MEA (Manajemen, Ekon. dan Akuntansi)*, vol. 7, no. 5, 2023.

[22] R. M. Taufiq, D. Toresa, dan S. Handayani, "Optimasi parameter Support Vector Machine dengan Particle Swarm Optimization untuk prediksi tunggakan iuran sekolah," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 75–84, 2024.

[23] M. F. Syahputra, F. M. Azizah, dan M. S. Lubis, "Optimasi parameter Support Vector Machine menggunakan algoritma genetika untuk meningkatkan prediksi pergerakan harga saham," *CORE: J. Comput. Res. Educ.*, vol. 3, no. 1, pp. 45–55, 2022.

[24] R. Nabilah, N. D. Arifatunisa, R. R. Fakhri, G. N. Fitriana, D. A. Firjatulloh, dan J. T. Nugraha, "Analisis kepuasan pelanggan terhadap layanan transportasi online (Grab)," in *Proc. Sem. Nasional Sist. Inform. Indonesia (SESINDO)*, 2024.

[25] R. S. Mongilala, J. A. F. Kalangi, dan O. F. C. Walangitan, "Kualitas pelayanan transportasi online (GrabBike) terhadap kepuasan pelanggan," *J. Ilm. Administrasi Bisnis*, vol. 1, no. 3, 2020.