User Sentiment Classification in Gojek: Machine Learning Model Comparison

Pipiana Anggrai Juni1), Salsa Deswina Raihani1), Natali Valentina Sutanto1), Alexander Agung Santoso Gunawan2,a), and Karli Eka Setiawan2)

1Data Science Program, Computer Science Department, Bina Nusantara University, Jakarta, Indonesia

2Computer Science Department, Bina Nusantara University, Jakarta, Indonesia

a) Corresponding author: [aagung@binus.edu](mailto:yos.sumantri@upnyk.ac.id)

**Abstract.** Sentiment represents a fundamental human emotion and comprehending them involves taking various emotions into account to identify problems and gain a deeper understanding of the scenario. Gojek's application has risen to prominence in the competitive ride-hailing and on-demand services sector, offering convenience and accessibility to a wide range of services. The research aims to utilize machine learning techniques, particularly the Random Forest, Support Vector Machine (SVM), and Decision Tree models, to assess their efficacy in sentiment classification, encountering a notable challenge in implementing these techniques within the realm of informal Indonesian text. Among the models employed in the study, the SVM model emerged as the top performer, showcasing an impressive 86% accuracy rate, surpassing both the Random Forest and Decision Tree models. These findings emphasize the potential efficacy of SVM in interpreting sentiments expressed in the informal Indonesian language, signaling its proficiency in this specific linguistic context.

**Keywords:** Sentiment Classification, Machine Learning, Gojek Apps, Indonesian Language, Text Mining.

# INTRODUCTION

In the era of massive digital transformation, mobile applications have become an integral part of everyday life, offering convenience and accessibility to a wide range of services. Gojek's application has emerged as a prominent player in the ride-hailing and on-demand services industry [1]. Alongside a highly competitive business landscape, the ability to harness sentiment analysis insights becomes a critical tool for staying competitive and ensuring long-term success, particularly for Gojek. This lies in its capacity to pinpoint strengths and weaknesses in its services, promptly address customer concerns, and implement data-driven improvements, thereby enhancing the user experience and fostering brand loyalty.

In contemporary society, individuals articulate their viewpoints and perspectives on various events and phenomena. Diverse technological tools are leveraged to scrutinize user sentiments [2]. The opinions and reactions of users play a pivotal role in shaping the trajectory of technological advancements or organizational progress. Consequently, Sentiment Analysis, a facet of natural language processing, emerges as a crucial mechanism for interpreting individuals' perspectives and rendering insightful opinion-based judgments [3]. The analysis of user sentiment employs a range of technologies, broadly categorized into machine learning and deep learning methods [4]. This research, situated within the domain of User Sentiment Classification in Gojek, investigates the application of Support Vector Machine (SVM), Random Forest, and Decision Tree models, representative components of machine learning, for an extensive comparative evaluation.

The research delves into the process of data collection, preprocessing, model development, and model testing, showcasing the applicability of SVM, Random Forest, and Decision Tree models in the context of Gojek's user sentiment classification. The final model predicts user sentiments as either positive, negative, or neutral regarding the Gojek application, aiming to elevate the overall user experience [5]. This research contributes to the convergence of natural language processing techniques and their practical implementation, offering valuable insights for organizations striving to enhance their services [6].

The research is systematically structured, commencing with section one providing a brief overview of the importance of sentiment analysis within the Gojek context and delves into previous studies concerning sentiment analysis in online transportation services, shedding light on the methodologies employed. Following this, section two meticulously outlines the Method, elaborating on stages involving dataset preparation, exploratory data analysis, and text preprocessing. This section also includes a discussion on modelling techniques using Support Vector Machine (SVM), Random Forest, and Decision Tree. Moving forward, section three systematically presents results and a comparative analysis of accuracy metrics. Lastly, section four succinctly summarizes the research methodology and key findings, underscoring the success of the implemented models. The chapter concludes by offering recommendations for future research and potential advancements in sentiment analysis.

The field of sentiment analysis within the realm of online transportation services has garnered considerable attention, with numerous researchers employing a variety of methodologies to gauge the sentiments expressed by users. These investigations, detailed below, have contributed to a better understanding of user sentiments and the quality of services within the online transportation industry. In the study conducted by Fitriyah et al. [7] Support Vector Machine (SVM) with linear and Radial Basis Function (RBF) kernels was employed. The manual sentiment labelling resulted in an impressive overall accuracy of 79.19%. This research focused on Twitter data and found that Gojek users tended to express positive sentiments during a specific period. Several studies implemented Naïve Bayes Classifiers such as the study conducted by Adilah et al. [8] with the Instagram platform collected data that achieved a noteworthy accuracy of 81.00%. Pratama et al. [9] with collected reviews from the Google Play Store comment section successfully achieved 84.44%. and lastly, Fahmi et al. [10] indicate a lower accuracy compared to other papers. Other studies conducted using a combination of Naïve Bayes and Support Vector Machine methods for classifying sentiments into positive, negative, and neutral categories including Petiwi et al. [11] and Hermanto et al. [12] along with the implementation of the SMOTE technique are successfully achieved substantial accuracies of 79.19% and 81.09% respectively. Another approach by Setiawan et al. [13] introduced an ensemble stacking approach using SVM with RBF and linear kernels, along with logistic regression has effectively achieved an accuracy rate of 88%. Also, a study conducted by Amalia et al. [14] employed multi-label classification and the Random Forest algorithm. Multi-label classification attained a testing accuracy of 76%, while the Random Forest algorithm showed excellent results with 94% accuracy for service category classification and 78% accuracy for sentiment classification. Lastly, a study by Arsarinia et al. [15] utilized Google Machine Learning to analyze social media data from Twitter, YouTube, and Google Search, achieving an impressive accuracy rate of 82.6%. These works collectively demonstrate the diverse methods and varying degrees of accuracy in sentiment analysis of online transportation services. Researchers have used methods such as SVM, Naïve Bayes, ensemble stacking, multi-label classification, and Google Machine Learning to gain insights into user sentiments, which can be valuable for improving service quality and understanding customer feedback. Several of these papers even compare several tests using different portions of data splitting ratio.

This research, comparing SVM, Random Forest, and Decision Tree models in user sentiment classification on Gojek, adds value to the existing literature by providing a comprehensive and in-depth perspective. Its implications are not limited to the context of Gojek but can contribute to the development of sentiment analysis methodologies in the broader context of the online transportation industry. The importance of this research is highlighted in the context of increasing competition in the online service sector, where a profound understanding of user sentiments can become a competitive advantage.

# METHODS

Throughout this research initiative, a systematic methodology, as illustrated in **FIGURE 1**, has been executed. The methodology commences with the preparation of the dataset, involving tasks such as loading data, labeling, conducting exploratory data analysis, text preprocessing, and adopting suitable text representation techniques. Subsequently, the research advances into the Modeling phase, concluding with the evaluation of model performance, utilizing benchmarks like accuracy. These benchmarks function as comparative metrics against preceding studies, delivering a thorough evaluation of the models employed in this research.

A diagram of a model

Description automatically generated

**FIGURE 1**. Methods

The research paper employed the "Data Aplikasi Gojek" dataset sourced from Kaggle for analysis, as seen in **FIGURE 2**. This dataset encompasses 20,000 reviews specifically gathered from the Gojek application accessible on the Google Play Store, compiled within the timeframe spanning January 2023 to March 2023.

A close up of words

Description automatically generated

**FIGURE 2**. Data contents (the data terms used are written in Indonesian)

The data labeling process involved the incorporation of a "sentiment" column (**FIGURE 3**), which relied on the existing "score" column within the dataset. The criteria utilized for classification were delineated as follows: assigning a negative sentiment value of -1 to scores ranging from 1 to 2, designating a neutral stance with a value of 0 for a score of 3, and categorizing scores between 4 and 5 as positive sentiment, represented by a value of 1.

A graph of a bar

Description automatically generated with medium confidence

**FIGURE 3**. Data distributions by sentiment

In the text preprocessing phase, various essential steps are undertaken to enhance the model's efficiency. These encompass the removal of whitespaces, punctuation, numbers, and special characters, alongside normalization, stop words elimination, stemming, and tokenization. The initial step involves converting all text to lowercase to maintain uniformity in word representation. Subsequently, systematic procedures are applied to eliminate extraneous elements such as whitespaces, punctuation marks, numbers, and special characters, contributing to the refinement of the text data. The data before and after text preprocessing is detailed in **TABLE 1.**

**TABLE 1**. Comparison of data before and after text preprocessing

|  |  |
| --- | --- |
| Before Lowercasing, Removing Whitespace, Punctuation, Numbers and Special Characters | After Lowercasing, Removing Whitespace, Punctuation, Numbers and Special Characters |
| MY GOPAY WALLET WAS BLOCKED WITH UNCLEAR REASONS I SUFFERED LOSSES WITH THE LOSS OF MY BALANCE REFUND MY MONEY | my gopay wallet was blocked with unclear reasons i suffered losses with the loss of my balance refund my money. |
| There are still stupid and foolish consumers who play orders with the excuse of squeezing lah and others remember karma applies to those of you who like to cancel orders. | there are still stupid and foolish consumers who play orders with the excuse of squeezing lah and others, remember karma applies to those of you who like to cancel orders. |
| I clicked on the desired destination point eh even after the driver received my order it became the destination point semoha debeloper fix again for the driver is ok | i clicked on the desired destination point eh even after the driver received my order it became the destination point semoha debeloper fix again for the driver is ok |
| want to transfer eh.... Upgrade... failed because of blurred ID card... there is no solution.....That's the NIK.....What | want to transfer eh upgrade fails first because ktp blur emang there is no other solution nik nya so what family card number do not be complicated |
| NO. FAMILY CARDS.... don't make it difficult.... I can't transfer because I tried to click gopay plus, it's been 2 days of providing data that hasn't been ACC'd yet... can't transfer between banks so | i can't transfer because i tried to click gopay plus it's been days of providing data that hasn't been accredited, so i can't transfer between banks. |

Following this, normalization approaches are employed to standardize the text, thereby ensuring consistency and bolstering the model's capability to interpret the information accurately. As exemplified in the **TABLE 2**, normalization involves transforming colloquial or abbreviated terms such as 'yg' to 'yang', 'gak' and 'gag' to 'enggak', and 'gt' to 'begitu'. This process aims to standardize and regularize diverse expressions, facilitating a more consistent and coherent dataset for analysis. The table below compares the data before and after normalization.

**TABLE 2**. Comparison of data before and after normalization

|  |  |
| --- | --- |
| Before Normalization | After Normalization |
| my wallet was blocked with unclear reasons I suffered losses with the loss of my balance refund my money | my wallet was blocked with unclear reasons I suffered losses with the loss of my balance refund my money |
| there are still stupid and foolish consumers who play orders with the excuse of squeezing lah and others remember karma berlkau for those of you who like to cancel orders | there are still stupid and foolish consumers who play orders with the excuse of squeezing lah and others remember karma berlkau for those of you who like to cancel orders |
| i clicked on the desired destination point eh even after the driver received my order even so different destination points hopefully the developer fixes it again for the driver anyway ok | i clicked on the desired destination point eh even after the driver received my order even so different destination points hopefully the developer fixes it again for the driver anyway ok |
| want to transfer eh upgrade fails first because the ktp is blurry there is no other solution nik what is the family card number do not be complicated | want to transfer eh upgrade fails first because the KTP is blurry there is no other solution NIK what is the family card number do not be complicated |
| i can't transfer because I tried to click gopay plus it's been days of providing data that hasn't been accredited and I can't transfer between banks. | i can't transfer because I tried to click gopay plus it's been days of providing data that hasn't been accredited, so I can't transfer between banks. |

To bolster the dataset's integrity, a method is executed to eradicate stopwords—words that frequently occur but potentially hold minimal influence on the outcome of sentiment analysis, as shown in **TABLE 3**. This approach aims to elevate the overall quality of the dataset by excluding commonly encountered words that might have a limited impact on the sentiment evaluation process. A comparison of the initial and after removing stopwords from the data is shown in the table below. The table depicted above showcases the elimination of specific terms like "saya" (I) identified as Conjunctions, "dan" (and) categorized as Prepositions, "di" (at/in/on) classified as Conjunctive Adverbs, "karena" (because) listed as Adverbs, along with "saja" (only), among others. Additionally, Adjectives have been meticulously excluded from the dataset.

**TABLE 3**. Comparison of data before and after removing stopwords

|  |  |
| --- | --- |
| Before Removing Stopword | Before Removing Stopword |
| my wallet was blocked with unclear reasons I suffered losses with the loss of my balance refund my money | ['wallet', 'gopay', 'blocked', 'reason', 'experienced', 'loss', 'missing', 'balance', 'return', 'money'] |
| there are still stupid and foolish consumers who play orders with the excuse of squeezing lah and others remember karma berlkau for those of you who like to cancel orders | ['consumer', 'stupid', follish, 'play', 'order', 'reason', 'squeeze', 'karma', 'apply', 'like', 'cancel', 'order'] |
| i clicked on the desired destination point eh even after the driver received my order even so different destination points hopefully the developer fixes it again for the driver anyway ok | ['click', 'point', 'destination', 'eh', 'driver', 'accept', 'order', 'different', 'point', 'destination', 'hopefully', 'developer', 'fix', 'driver', 'anyway', 'ok'] |
| want to transfer eh upgrade fails first because the ktp is blurry there is no other solution nik what is the family card number do not be complicated | ['transfer', 'eh', 'upgrade', 'fail', 'mulu', 'because', 'ktp', 'blur', 'solution', 'nik', 'nya', 'no', 'card', 'family', 'complicate'] |
| i can't transfer because I tried to click gopay plus it's been days of providing data that hasn't been accredited and I can't transfer between banks. | ['transfer', 'because', 'because', 'coba', 'klik', 'gopay', 'plus', 'mengasih', 'data', 'acc', 'transfer', 'bank'] |

Following the normalization phase, the subsequent implementation of stemming assumes considerable importance, shown in **TABLE 4**. Stemming, a method utilized to condense words into their fundamental or root forms, assumes a critical role in unifying diverse word variations. This approach facilitates feature reduction and simplification of subsequent analytical processes, as exemplified in the table provided below.

**TABLE 4**. Comparison of data before and after stemming

|  |  |
| --- | --- |
| Before Stemming | After Stemming |
| ['wallet', 'gopay', 'blocked', 'reason', 'experienced', 'loss', 'missing', 'balance', 'return', 'money'] | gopay wallet blocking reason for loss lost saldi return money |
| ['consumer', 'stupid', follish, 'play', 'order', 'reason', 'squeeze', 'karma', 'apply', 'like', 'cancel', 'order'] | stupid consumers dumb enough to order karma excuses sell like cancel orders |
| ['click', 'point', 'destination', 'eh', 'driver', 'accept', 'order', 'different', 'point', 'destination', 'hopefully', 'developer', 'fix', 'driver', 'anyway', 'ok'] | click the destination point eh driver receive a message different destination point hopefully good developer driver anyway okay |
| ['transfer', 'eh', 'upgrade', 'fail', 'mulu', 'because', 'ktp', 'blur', 'solution', 'nik', 'nya', 'no', 'card', 'family', 'complicate'] | tranfer eh upgrade failed mulu because ktp blur nik solution family card difficult |
| ['transfer', 'because', 'because', 'try', 'klik', 'gopay', 'plus', 'give', 'data', 'acc', 'transfer', 'bank'] | transfer because I tried to click gopay plus give bank transfer acc data |

Alongside stemming, the application of tokenization is integral, dividing the text into smaller units like words or phrases. This fragmentation greatly eases the conversion of textual information into numerical vectors, a crucial step for machine learning model input. From the presented **TABLE 5**, the initial state reveals raw textual data in its original, unstructured form, where sentences and phrases are displayed as cohesive strings. Following the tokenization process, a transformation occurs, rendering the table with a structured representation of the same textual information. Each sentence or phrase is disassembled into individual units, commonly referred to as tokens, which typically correspond to words or sub words. This structured format provides a basis for more detailed analysis and processing, enabling a nuanced comprehension of the textual content.

**TABLE 5**. Comparison of data before and after tokenization

|  |  |
| --- | --- |
| Before Tokenzination | After Tokenzination |
| gopay wallet blocking reason for loss lost saldi return money | [455, 8, 106, 174, 530, 209, 66, 30, 195, 69] |
| stupid consumers dumb enough to order karma excuses sell like cancel orders | [243, 721, 329, 299, 18, 174, 541, 2027, 403, 56, 1535, 18] |
| click the destination point eh driver receive a message different destination point hopefully good developer driver anyway okay | [290, 142, 122, 204, 4, 36, 17, 155, 142, 122, 57, 649, 29, 4, 35, 11] |
| transer eh upgrade failed mulu because ktp blur nik solution family card difficult | [68, 204, 134, 157, 116, 3761, 203, 1540, 194, 1725, 2, 146, 341, 334, 166] |
| transfer because I tried to click gopay plus give bank transfer acc data | [68, 245, 245, 80, 290, 8, 128, 265, 227, 725, 68, 109] |

Exploratory Data Analysis (EDA) helps understand dataset characteristics and extract insights before model building. It reveals patterns, focusing on sentiment analysis of user comments with sentiment labels. Initial analysis uses basic statistics to examine sentiment distribution and word frequency to identify key terms. **FIGURE 4.** shows the top 20 most frequent words, highlighting prevalent themes in Gojek user reviews, such as driver performance, app satisfaction, GoPay functionality, pricing issues, promotions, and service speed.

A graph of a number of words

Description automatically generated

**FIGURE 4**. Top 20 most frequently occurring words (the data terms used are written in Indonesian)

Topic modeling with Latent Dirichlet Allocation (LDA) shows that key themes include user satisfaction with drivers and service fee discounts, indicating a link between satisfaction and fee reductions, as seen in **TABLE 6.** It also highlights the need to address issues with discount vouchers and application-related challenges, as well as technical problems affecting application ease. These findings emphasize the importance of resolving technical issues to improve user experience.

**TABLE 6**. Keywords and the description from LDA

|  |  |  |
| --- | --- | --- |
| Topic | Keywords | Description |
| 1 | 'fast', 'friendly', 'driver', 'cut', 'driver', 'benefit', 'polite', 'its', 'send', 'bug’ | Speed of service, driver’s politenesss and discounts or cuts in Gojek services. |
| 2 | 'steady', 'ok', 'cool', 'interesting', 'satisfied', 'cash', 'staple', 'satisfied', 'really', 'nice' | Positive responses or user satisfaction with the service. |
| 3 | 'gopay', 'login', 'gojek', 'its', 'balance', 'account', 'okay', 'app', 'use', 'yes' | Usage of GoPay, balance, and the Gojek application. |
| 4 | 'good', 'pay', 'I', 'gopaylater', 'date', 'use', 'paylater', 'bill', 'later', 'installment' | User experiences using payment features in the app. |
| 5 | 'help', 'love', 'receive', 'app', 'gojek', 'hope', 'like', 'need', 'really', 'go' | Positive responses or user satisfaction with the service. |
| 6 | 'use', 'promo', 'gojek', 'app', 'nya', 'voucher', 'go', 'nya', 'error', 'better' | Use of promos, vouchers, and technical issues like errors in the application. |
| 7 | 'driver', 'gojek', 'its', 'order', 'message', 'help', 'love', 'star', 'if', 'app' | Interactions with drivers, placing orders, and feedback regarding Gojek’s driver service. |
| 8 | 'serve', 'apk', 'safe', 'bad', 'its', 'convenient', 'address', 'app', 'fast', 'app' | Concerns service quality, security, and the user experience of using the Gojek application. |
| 9 | 'gojek', 'easy', 'app', 'error', 'thank you', 'road', 'happy', 'steady', 'use', 'best' | Ease of using the app and technical issues. |
| 10 | 'good', 'cost', 'expensive', 'its', 'application', 'price', 'really', 'thousand', 'eat', 'serve' | User assesments of picing and service cost provided by Gojek |

Regarding text representation phase, researchers used the TF-IDF method to convert each review into a numerical vector. The 'TfidfVectorizer' from scikit-learn calculates word frequency (Term Frequency) and significance across documents (Inverse Document Frequency), resulting in numeric vectors that enhance machine learning model performance on text data. In the modeling phase, three machine learning algorithms—Support Vector Machine (SVM), Random Forest, and Decision Tree—were used to classify user sentiment on the Gojek platform. SVM, employing a linear kernel, is effective for sentiment analysis by finding the optimal hyperplane for class separation. Random Forest, an ensemble method that builds multiple decision trees, enhances classification performance and robustness [16]. Decision Trees use a tree-like structure to make decisions based on feature values and recursively split the data [17]. The study compares these models based on accuracy, highlighting their strengths and weaknesses in user sentiment classification. The evaluation of SVM, Random Forest, and Decision Tree models is based on accuracy, which measures the ratio of correctly predicted instances to the total number of instances. Higher accuracy indicates better performance, reflecting the models' effectiveness in capturing data patterns and making accurate predictions.

# RESULTS AND DISCUSSION

Among the analyzed models, the Support Vector Machine (SVM) demonstrated slightly superior performance, achieving an impressive accuracy rate of 86%. This suggests that SVM is particularly adept at handling the complexity of sentiment analysis, where the goal is to classify sentiments accurately across various contexts. The model's effectiveness can be attributed to its ability to manage intricate decision boundaries with precision. Additionally, SVM's inherent capability to minimize overfitting by maximizing the margin between classes makes it more robust in predicting unseen data, a key factor in its strong performance. In comparison, the Random Forest model also performed robustly, attaining an accuracy rate of 85%. This model benefits from an ensemble learning approach with multiple decision trees, effectively capturing diverse sentiment patterns on the Gojek platform. Despite a slightly lower accuracy of 82% compared to SVM, it remains effective and interpretable. The Decision Tree model, though prone to overfitting, offers valuable insights into sentiment classification. In comparison as presented in **TABLE 7**, the SVM model in this study, with accuracy slightly below the 88% reported by Setiawan et al. [13], outperforms the 79.19% accuracy from Fitriyah et al. [7], showing strong tuning and effectiveness in sentiment analysis for online transportation services.

**TABLE 7**. Model accuracy comparison

|  |  |
| --- | --- |
| **Model** | **Average accuracy** |
| **Support Vector Machine (SVM)** | 86% |
| **Random Forest** | 85% |
| **Decision Tree** | 82% |

# CONCLUSIONS

In conclusion, this research demonstrates the efficacy of Support Vector Machines (SVM), Random Forest, and Decision Tree models in sentiment analysis, particularly within the context of Gojek's user reviews. The SVM model, with its accuracy rate of 86%, emerged as the most effective, while the Random Forest and Decision Tree models also delivered strong performances, achieving accuracy rates of 85% and 82%, respectively. These results not only surpass benchmarks established in related works but also underscore the potential of these models in real-world applications. The findings of this study have significant practical implications for Gojek and similar online transportation platforms. Implementing an SVM-based sentiment classification system could enhance Gojek's ability to monitor and respond to customer feedback in real time. By accurately categorizing user sentiments, Gojek can gain deeper insights into customer satisfaction, identify areas for service improvement, and tailor marketing strategies to better meet customer needs. Furthermore, integrating sentiment analysis into customer service workflows could help prioritize issues that require immediate attention, ultimately leading to an improved user experience. Looking ahead, future research could explore the integration of advanced text representation techniques such as word embeddings, contextual embeddings, or transformer-based models like BERT or GPT. These approaches have the potential to further improve the accuracy and robustness of sentiment classification systems. Additionally, the exploration of deep learning techniques, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Expanding this research to include different linguistic and cultural contexts could also yield insights into the adaptability and generalizability of these models across diverse datasets, contributing to the broader field of sentiment analysis in online services.

# Acknowledgments

The authors of this journal article are Pipiana, Natali, and Salsa, based on research titled "User Sentiment Classification in Gojek: Machine Learning Model Comparison," funded by Binus University's Faculty of School of Computer Science. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the funding agency. This journal would not have been possible without the professional support and assistance provided by the esteemed reviewers. Special acknowledgment is extended to:

1. Dr. Ir. Alexander Agung Santoso Gunawan S.Si., M.T., M.Sc., Bina Nusantara University, Indonesia, for his guidance and supervision.
2. Karli Eka Setiawan, S.Si., M.Kom., Bina Nusantara University, Indonesia, for his guidance and review.

# References

1. Wahyu Handani, S., et al., *Sentiment Analysis for Go-Jek on Google Play Store.* Journal of Physics: Conference Series, 2019. **1196**(1): p. 012032.

2. Srinivas, A.C.M.V., et al., *Sentiment Analysis using Neural Network and LSTM.* IOP Conference Series: Materials Science and Engineering, 2021. **1074**(1): p. 012007.

3. Khan, L., et al., *Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media.* Applied Sciences, 2022. **12**(5): p. 2694.

4. Tan, K.L., C.P. Lee, and K.M. Lim, *A survey of sentiment analysis: Approaches, datasets, and future research.* Applied Sciences, 2023. **13**(7): p. 4550.

5. Djatmiko, F., R. Ferdiana, and M. Faris. *A Review of Sentiment Analysis for Non-English Language*. in *2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*. 2019.

6. Suriah, G.K., et al. *Analysis of Performance Long Short-Term Memory-Convolutional Neural Network (LSTM-CNN) for Lifelong Learning on Indonesian Sentiment Analysis*. in *Advanced Intelligent Systems for Sustainable Development (AI2SD’2020)*. 2022. Cham: Springer International Publishing.

7. Fitriyah, N., B. Warsito, and I.M. Di Asih, *Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM).* Jurnal Gaussian, 2020. **9**(3): p. 376-390.

8. Tika Adilah, M., et al., *Sentiment Analysis of Online Transportation Service using the Naïve Bayes Methods.* Journal of Physics: Conference Series, 2020. **1641**(1): p. 012093.

9. Arif Pratama Budiman, A., *Implementasi Metode Learning Vector Quantization (LVQ) UntukSentimen Analisis Terhadap Aplikasi Go-Jek Pada Playstore.* Implementasi Metode Learning Vector Quantization (LVQ) UntukSentimen Analisis Terhadap Aplikasi Go-Jek Pada Playstore, 2022. **5**(3): p. 364-373.

10. Fahmi, M., Y. Yuningsih, and A. Puspita, *Sentiment Analysis Of Online Gojek Transportation Services On Twitter Using The Naïve Bayes Method.* JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer), 2023. **8**(2): p. 90-96.

11. Petiwi, M.I., A. Triayudi, and I.D. Sholihati, *Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine.* Jurnal Media Informatika Budidarma, 2022. **6**(1): p. 542-550.

12. Hermanto, et al., *Gojek and Grab User Sentiment Analysis on Google Play Using Naive Bayes Algorithm And Support Vector Machine Based Smote Technique.* Journal of Physics: Conference Series, 2020. **1641**(1): p. 012102.

13. Setiawan, Y., J. Jondri, and W. Astuti, *Twitter Sentiment Analysis on Online Transportation in Indonesia Using Ensemble Stacking.* JURNAL MEDIA INFORMATIKA BUDIDARMA, 2022. **6**(3): p. 1452-1458.

14. Amalia, A., D. Gunawan, and K. Nasution, *Sentiment analysis of GO-JEK services quality using Multi-Label Classification.* Journal of Physics: Conference Series, 2021. **1830**(1): p. 012003.

15. Arsarinia, D.A.P.S., I.K.G.D. Putraa, and N.K. Dwi, *Public Sentiment Analysis of Online Transportation in Indonesia through Social Media Using Google Machine Learning.* Jurnal Ilmiah Merpati, 2021. **9**: p. 153-164.

16. Biau, G., *Analysis of a random forests model.* The Journal of Machine Learning Research, 2012. **13**(1): p. 1063-1095.

17. Meenu, S.G., *Sentiment Analysis using Decision Tree.* International Journal of Electronics Engineering, 2019. **11**(1): p. 965-970.