Firman Noor Hasan

Fathurrohman, Afandi, Wahyuningtyas, Nugroho, Hasan - cek.01

IJID, 2025 - Fathurrohman, Afandi, Wahyuningtyas, Nugroho, Hasan

Project Paper

Universitas Muhammadiyah Prof. Dr. Hamka

Document Details

Submission ID

trn:oid:::1:3177988252

Submission Date

Mar 10, 2025, 6:47 AM GMT+7

Download Date

Mar 10, 2025, 6:49 AM GMT+7

File Name

Jurnal_SPX_-_cek.01.docx

File Size

609.2 KB

18 Pages

12,406 Words

78,164 Characters

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Top Sources

8% Internet sources
7% Publications
2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

8% Internet sources
7% Publications
2% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
repository.uhamka.ac.id		<1%
2	Internet	
www.ojs.uma.ac.id		<1%
3	Internet	
discovery.researcher.life		<1%
4	Internet	
www.jetir.org		<1%
5	Publication	
T. Mariprasath, Kumar Reddy Cheepati, Marco Rivera. "Practical Guide to Machin...		<1%
6	Internet	
ebin.pub		<1%
7	Internet	
www.mdpi.com		<1%
8	Internet	
www.ijfmr.com		<1%
9	Internet	
ai.jmir.org		<1%
10	Internet	
portalcris.vdu.lt		<1%
11	Internet	
journal.ipm2kpe.or.id		<1%

12 Publication

H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He... <1%

13 Publication

H.L. Gururaj, Francesco Flammini, J. Shreyas. "Data Science & Exploration in Artifi... <1%

14 Student papers

Queen Mary and Westfield College <1%

15 Internet

www.geeksforgeeks.org <1%

16 Internet

docs.google.com <1%

17 Internet

ejurnal.seminar-id.com <1%

18 Publication

Achmad Warsito Sujatmiko, Anik Vega Vitianingsih, Slamet Kacung, Dwi Cahyono,... <1%

19 Publication

Sai Kiran Oruganti, Dimitrios A Karras, Srinesh Singh Thakur, Janapati Krishna Ch... <1%

20 Internet

api-kwasuspace.kwasu.edu.ng <1%

21 Internet

heca-analitika.com <1%

22 Publication

Araújo, Tatiana. "Informed Retail Product Similarity", Universidade do Porto (Port... <1%

23 Internet

jutif.if.unsoed.ac.id <1%

24 Internet

mufad.org.tr <1%

25 Internet

robots.net <1%

26 Publication

Melda Betaria Sitepu, Ibnu Rasyid Munthe, Syaiful Zuhri Harahap. "Implementati... <1%

27 Internet

dspace.daffodilvarsity.edu.bd:8080 <1%

28 Internet

jurnal.bimaberilmu.com <1%

29 Internet

ojs.uni-miskolc.hu <1%

30 Publication

J Jumadi, D S Maylawati, L D Pratiwi, M A Ramdhani. "Comparison of Nazief-Adria... <1%

31 Internet

infectagentscancer.biomedcentral.com <1%

32 Internet

en.istanbulkongresi.org <1%

33 Internet

jurnal.iaii.or.id <1%

34 Internet

www.coursehero.com <1%

35 Internet

www.ijirmf.com <1%

36 Internet

www.scaler.com <1%

37 Internet

d197for5662m48.cloudfront.net <1%

38 Internet

deepai.org <1%

39 Internet

dspace.univ-ouargla.dz <1%

40 Internet

files.eric.ed.gov <1%

41 Internet

npws.nsw.gov.au <1%

42 Internet

www.biorxiv.org <1%

43 Internet

archive.umsida.ac.id <1%

44 Internet

eprints.ucm.es <1%

45 Internet

link.springer.com <1%

46 Internet

ndltd.ncl.edu.tw <1%

47 Internet

www.youscribe.com <1%

48 Internet

bilselkongreleri.com <1%

49 Publication

da Cunha Alcaide, Diogo. "Predicting Lapse Rate in Life Insurance: An Exploration... <1%

50 Internet

doczz.biz.tr <1%

51 Internet

hdl.handle.net <1%

52 Internet

irjet.net <1%

53 Internet

kabinetrakyat.com <1%

54 Internet

researchbank.swinburne.edu.au <1%

55 Internet

www.giikorea.co.kr <1%

56 Publication

Ajay Kumar, Shashank Sheshar Singh, Kuldeep Singh, Bhaskar Biswas. "Link predi... <1%

57 Publication

E. Navas. "An Objective and Subjective Study of the Role of Semantics and Prosodi... <1%

58 Publication

Nathanael Yudhistira Pradipta, Hari Soetanto. "Sentiment Classification of Gener... <1%

59 Internet

docplayer.com.br <1%

60 Internet

fastercapital.com <1%

61 Internet

iris.cnr.it <1%

62 Internet

journals.plos.org <1%

63 Internet

onlinemanuals.txdot.gov <1%

64 Internet

repositorio.ulima.edu.pe <1%

65 Internet

sciendo.com <1%

66 Internet

sdn.ifmo.ru <1%

67 Internet

www.dovepress.com <1%

68 Publication

Girard, Fantin. "Modelisation Statistique Des Structures Anatomiques De La retin... <1%

69 Publication

A. Malini, Surbhi Bhatia Khan, S. Kayalvizhi, Mohammed Saraee. "Advances in Co... <1%

70 Publication

Fabrício Almeida do Carmo, Jorge Luiz Figueira da Silva Junior, Rafael Geraldeli Ro... <1%

71 Publication

Kutub Thakur, Helen G. Barker, Al-Sakib Khan Pathan. "Artificial Intelligence and ... <1%

72 Internet

towardsdatascience.com <1%

73 Publication

Çimen, Seyda Selen. "A Case Study on Pre-Service English Language Teachersâ€Ÿ ... <1%

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

Sentiment Analysis of Shopee Xpress Reviews Using Support Vector Machine and Logistic Regression

Sewin Fathurrohman [1]
Department of Information Systems
Universitas Gunadarma
Depok, Indonesia
sewin.fathur@gmail.com

Irfan Ricky Afandi [2]
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
richky99@gmail.com

Irma Wahyuningtyas [3]
Department of Mathematics Education
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
3melatiputih@gmail.com

Azis Styo Nugroho [4]
Department of Automation Engineering Technology
Universitas Negeri Jakarta
Jakarta, Indonesia
fled0211@gmail.com

Firman Noor Hasan [5*]
Department of Informatics Engineering
Universitas Muhammadiyah Prof. Dr. Hamka
Jakarta, Indonesia
firman.noorhasan@uhamka.ac.id

Article History
Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy
Published month, yyyy

Abstract*—The rapid expansion of e-commerce has increased the demand for efficient logistics services, making customer feedback analysis crucial for service improvement. Shopee Xpress receives a large volume of user reviews, making manual sentiment evaluation infeasible. This study applied sentiment analysis to classify Shopee Xpress customer reviews into positive, neutral, and negative sentiments using Support Vector Machine (SVM) and Logistic Regression. Customer reviews were collected from X (formerly Twitter) and Google Play Store. Data preprocessing techniques, including cleansing, casefolding, normalization, tokenization, stopword removal, and stemming. Sentiment classification performance was evaluated using accuracy, precision, recall, and F1-score. The results indicated that Logistic Regression achieved slightly higher accuracy (82%) than SVM (81%). Negative reviews predominantly highlighted delivery delays and service quality issues, whereas positive reviews emphasized affordability and efficiency. Both models demonstrated a low misclassification rate, with Logistic Regression excelling in detecting negative sentiments and SVM performing better in identifying neutral sentiments. These findings provide valuable insights for Shopee Xpress to optimize its logistics services by addressing key customer concerns. Future research could explore deep learning techniques to enhance classification accuracy and incorporate broader data sources for a more comprehensive analysis.

***Keywords*—logistic regression; machine learning; sentiment analysis; shopee xpress; support vector machine**

DOI: IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

1 INTRODUCTION

The development of e-commerce in Indonesia has experienced rapid growth in recent years, along with an increase in digital transactions and changes in consumer shopping behavior [1]. One of the largest e-commerce platforms in Indonesia is Shopee, which continues to innovate in providing the best services for its users [2]. To ensure efficiency in the delivery process, Shopee has developed its own logistics service known as Shopee Xpress [3]. This logistics service aims to enhance delivery speed and reliability, ultimately improving customer satisfaction [4].

In the logistics industry, competition among service providers is becoming increasingly intense, with each company striving to attract and retain customers [5]. One of the key factors influencing customer loyalty to a logistics service is their level of satisfaction with the service provided [6]. Customer satisfaction is often reflected in reviews or feedback shared on various digital platforms, such as social media and app stores [7] [8]. These reviews vary in content and sentiment [9], but they are often unstructured making it difficult to extract meaningful insights [10].

In today's digital era, customer reviews can become valuable assets for companies if properly managed and analyzed [11]. One effective approach to processing customer reviews is sentiment analysis [12], which classifies customer opinions into positive, negative, or neutral categories [13]. Sentiment analysis helps businesses understand customer perceptions of their services and serves as a basis for strategic decision-making [14]. Therefore, this study aims to conduct sentiment analysis on customer reviews of Shopee Xpress using machine learning methods. The selection of appropriate methods and algorithms is crucial to ensure that the extracted information is accurate and reliable [15].

This study employs two machine learning algorithms: Support Vector Machine (SVM) and Logistic Regression. SVM is a classification method that can handle both linear and nonlinear data by mapping nonlinear data into a higher-dimensional space to facilitate classification [16]. The working principle of SVM is to find the optimal hyperplane that separates different classes with the maximum possible margin [17]. Due to its flexibility, SVM has been widely applied in text-based sentiment analysis and has demonstrated high accuracy in classification tasks [18]. However, one limitation of SVM is that when features are highly similar or redundant, it can significantly impact the model's accuracy [19]. Logistic Regression is a statistical model used to predict the probability of binary (two-class) or multi-class outcomes based on given features [20]. This algorithm works by establishing a relationship between input variables and output probabilities through the sigmoid function [21]. The simplicity and effectiveness of Logistic Regression make it a widely used method, especially when combined with feature extraction techniques such as TF-IDF [22]. However, classical Logistic Regression tends to perform poorly when dealing with datasets where the number of features exceeds the number of samples [23].

Several previous studies have implemented sentiment analysis using similar algorithms in different contexts. For example, one study applied SVM to analyze customer reviews of the MyPertamina application on Google Play Store, achieving an accuracy of 85.31% in classifying sentiments into positive and negative categories [24]. Another study used Logistic Regression for sentiment analysis of online transportation services on the X platform, classifying sentiments into three categories (positive, neutral, and negative) and achieving an accuracy of 85% [25]. Unlike these studies, the present research integrates both algorithms and utilizes data from two sources it's from platform X and Google Play Store to provide a more comprehensive analysis of customer sentiment towards Shopee Xpress.

This study aims to provide deeper insights into customer sentiment regarding Shopee Xpress, as well as evaluate the effectiveness of SVM and Logistic Regression in sentiment classification. The findings of this research are expected to serve as a reference for Shopee Xpress and other logistics service providers in improving their service quality through data-driven analysis.

2 METHOD

This section provides an overview of the methodology applied in this study. The research was conducted using Google Colab and Python as the primary computational tools. Google Colab was chosen due to its accessibility, support for cloud-based GPU acceleration, and compatibility with a wide range of data mining libraries [26]. These features allow for efficient data processing and model training without requiring extensive local computational resources [27].

The methodology employed in this study follows a structured workflow to ensure the robustness and reproducibility of sentiment analysis. As illustrated in Fig. 1, the research process consists of several key stages in which each stage plays a crucial role in ensuring the accuracy and reliability of the sentiment classification results.

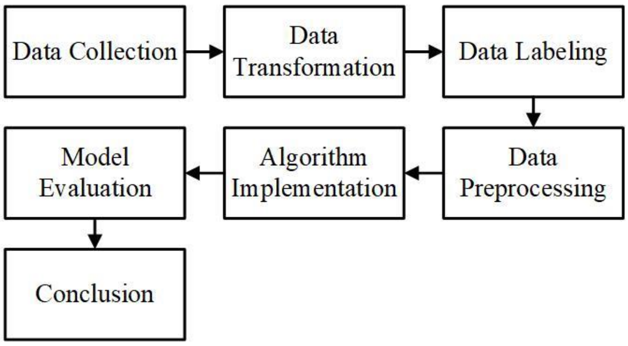

Figure 1. Workflow diagram of research

The research methodology follows a structured workflow, as illustrated in Fig. 1, which consists of seven main stages:

- *Data Collection:* Customer reviews related to Shopee Xpress are gathered from X and Google Play Store using web scraping techniques.
- *Data Transformation:* The collected datasets are preprocessed to ensure consistency, including attribute selection and format standardization.
- *Data Labeling:* Sentiment annotations (positive, negative, neutral) are assigned to a subset of the dataset by researcher.
- *Data Preprocessing:* Several text preprocessing techniques, such as cleansing, case folding, normalization, tokenization, stopword removal, and stemming, are applied to refine the textual data.
- *Algorithm Implementation:* The dataset is transformed into numerical representations using TF-IDF before being processed by machine learning models, namely Support Vector Machine (SVM) and Logistic Regression.
- *Model Evaluation:* The classification performance is assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score, based on the confusion matrix.
- *Conclusion:* The final analysis interprets the sentiment distribution and evaluates the model's effectiveness in classifying customer sentiments regarding Shopee Xpress services.

The following sections provide a comprehensive explanation of each step.

2.1 Data Collection

This research utilizes customer reviews from X and Google Play Store as the primary data sources. The collected reviews specifically mention Shopee Xpress, allowing for an in-depth sentiment analysis of customer experiences with the expedition service.

2.1.1 *The Search Keywords and Time Frame:* To retrieve relevant data, the keywords "shopee xpress" and "spx" were used as search queries. The dataset spans from January 1, 2024, to March 3, 2025, ensuring a comprehensive collection of recent user opinions.

2.1.2 *Data Acquisition Method:* The data was extracted using web scraping techniques, leveraging the API services provided by both platforms. The scraping process involved retrieving publicly available customer reviews based on predefined search criteria. To maintain data integrity and avoid redundancy, duplicate entries were removed after extraction.

2.1.3 *Dataset Overview:* The final dataset consists of 3100 reviews with 542 records from X and 2558 records from the Google Play Store, each containing multiple attributes. X Dataset had attributes include conversation_id_str, created_at, favorite_count, full_text, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str, and username. Google Play Store Dataset had attributes include reviewId, userName, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, and appVersion.

2.2 Data Transformation

The datasets obtained from X and Google Play Store possess distinct attribute structures, necessitating a transformation process to ensure uniformity and compatibility. This transformation involves attribute selection, date normalization, and dataset integration to facilitate a coherent sentiment analysis.

2.2.1 *Attribute Selection and Standardization:* Given the differences in attribute naming and structure, the following adjustments were applied:

- *X dataset:* The selection of attributes used in this research are “created_at” and “full_text”. The “created_at” attribute, originally in a timestamp format, was converted into a “date” field. Meanwhile, the “full_text” attribute was extracted and renamed as “review” to align with the dataset from Google Play Store.
- *Google Play Store dataset:* The selection of attributes used in this research are “at” and “content”. The “at” attribute, representing the timestamp of each review, was transformed into a “date” field. Similarly, the “content” attribute was renamed as “review”.

2.2.2 *Date Format Normalization:* To ensure consistency in temporal data representation, all dates from X and Google Play Store dataset were standardized to the YYYY-MM-DD format. This normalization ensures that reviews from both platforms are temporally aligned, enabling chronological sorting and trend analysis.

2.2.3 *Dataset Integration and Sorting:* After transforming and standardizing the attributes, the datasets from X and Google Play Store were merged into a single structured dataset. The consolidated dataset was subsequently sorted in descending order, with the most recent reviews appearing first. The final dataset consists of 3,100 reviews, ensuring a comprehensive and temporally organized corpus for sentiment analysis.

DOI: IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

The transformation process ensures consistency across datasets and facilitates seamless preprocessing for subsequent sentiment analysis [28]. A subset of the transformed dataset is presented in Table 1, illustrating the standardized format after the transformation process.

Table 1. Sample Transformed Dataset

Date	Review
2024-02-21	@ShopeeXpres @ShopeeID Gokil sih pakai ekspedisi spx 📦✨ paket nyampe cpt dari estimasi, kurirnya juga ramah. rekomendasi banget dan bakal pakai terus buat blnja online di shopee! 👍 #ShopeeXpress #Rekomendasi
2024-03-07	@ShopeeCare @ShopeeID Pakai ShopeeXpress buat pertama kali, lumayan sih. Kadang cepat, kadang telat pket sampenya. Kurirnya baik, tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki.
2024-06-10	Tlng dong kurirnya lebih hati-hati! Barang yang aku pesan sampe pecah karena dilempar atau gimana gitu, sama ada bekas robekannya juga di kemasan paketnya. Jujur kaya kecewa banget sama plyanan dr ShopeeXpress, kalua gitu kayanya mending pake ekspedisi lain yg lebih aman.
2024-10-19	Pernah coba buat retur barang pas pakai jasa ekspedissi ShopeeXpress. Prosesnya kagak yang terlalu ribet bangt sih, tapi lumayan lama juga nunggunya. Paketnya baru diambil kurir setelah 2 hari, terus butuh 5 hari lagi buat sampai ke seller. Jadi kalau butuh retur cepat bisa cari opsi lain.
2024-12-04	@ShopeeID Sumpah kecewa bgt! pake jasa ekspedisi spx Paket udah 4 hari stuck di gudang, trs status brang kagak update sama sekali di aplikasinya, trs mana CS susah dihubungi pula. Tolong dibantu donk @ShopeeCare, ini barang urgent! lambat banget pengirimannya ☹ #ShopeeXpress #Lambat
2025-01-22	Awalnya ragu pakai jasa ekspedisi ShopeeXpres, tapi ternyata lumayan oke banget sih menurut gw. Gw pesen paket dari Jakarta ke Bandung cuma 5 hari dah sampe cuy, update tracking juga jelas di aplkasi Shopeenya. apalagi ada promo gratis ongkirnya dan paket sampe dengan aman. Semoga nggak turun kualitas pengirimannya ke depannya ya shopee! kasih bintang 5 ☆☆☆☆☆ dulu deh hehehehe

2.3 Data Labeling

The labeling process was conducted manually by a team of five researchers, ensuring a thorough and systematic classification of sentiment. Each review was categorized into one of three sentiment classes: positive, negative, or neutral.

2.3.1 *Classification and Resolution:* Given the subjective nature of sentiment analysis, discrepancies in label assignment among researchers were inevitable. In cases of disagreement, two resolution strategies were employed. First resolution is discussion to achieve a mutual agreement on the most appropriate sentiment label. Second resolution is majority voting wherein the sentiment label agreed upon by at least three out of five researchers was assigned as the final label. This approach ensured the reliability and consistency of the labeled dataset, reducing potential biases in the classification process.

2.3.2 *Dataset Splitting for Model Training and Testing:* To facilitate machine learning model development, the dataset was randomly split into training and testing subsets. There are 80% of the data (2480 reviews) had been manually labeled by researchers, was allocated for training to enable the model to learn patterns effectively and 20% of the data (620 reviews) consisted of unlabeled data, was reserved for testing to objectively evaluate model performance. This 80:20 split was chosen to maintain a balance between model generalization and reliable performance assessment [29], ensuring that the model does not overfit the training data while being adequately tested on unseen samples [30].

The labeled dataset serves as the foundation for training and evaluating machine learning models, ensuring the robustness and validity of the sentiment classification process [31]. A sample of the labeled data is presented in Table 2, showcasing the sentiment classification assigned by the researchers.

Table 2. Sample of Labeled Data

Date	Review	Label
2024-02-21	@ShopeeXpres @ShopeeID Gokil sih pakai ekspedisi spx 📦✨ paket nyampe cpt dari estimasi, kurirnya juga ramah. rekomendasi banget dan bakal pakai terus buat blnja online di shopee! 👍 #ShopeeXpress #Rekomendasi	positive
2024-03-07	@ShopeeCare @ShopeeID Pakai ShopeeXpress buat pertama kali, lumayan sih. Kadang cepat, kadang telat pket sampenya. Kurirnya baik, tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki.	neutral
2024-06-10	Tlng dong kurirnya lebih hati-hati! Barang yang aku pesan sampe pecah karena dilempar atau gimana gitu, sama ada bekas robekannya juga di kemasan paketnya. Jujur kaya kecewa banget sama plyanan dr ShopeeXpress, kalua gitu kayanya mending pake ekspedisi lain yg lebih aman.	negative
2024-10-19	Pernah coba buat retur barang pas pakai jasa ekspedissi ShopeeXpress. Prosesnya kagak yang terlalu ribet bangt sih, tapi lumayan lama juga nunggunya. Paketnya baru diambil kurir setelah 2 hari, terus butuh 5 hari lagi buat sampai ke seller. Jadi kalau butuh retur cepat bisa cari opsi lain.	neutral
2024-12-04	@ShopeeID Sumpah kecewa bgt! pake jasa ekspedisi spx Paket udah 4 hari stuck di gudang, trs status brang kagak update sama sekali di aplikasinya, trs mana CS susah dihubungi pula. Tolong dibantu donk @ShopeeCare, ini barang urgent! lambat banget pengirimannya ☹ #ShopeeXpress #Lambat	negative
2025-01-22	Awalnya ragu pakai jasa ekspedisi ShopeeXpres, tapi ternyata lumayan oke banget sih menurut gw. Gw pesen paket dari Jakarta ke Bandung cuma 5 hari dah sampe cuy, update tracking juga jelas di aplkasi Shopeenya. apalagi ada promo gratis ongkirnya dan paket sampe dengan aman. Semoga nggak turun kualitas pengirimannya ke depannya ya shopee! kasih bintang 5 dulu deh hehehehe ☆☆☆☆☆	positive

2.4 Data Preprocessing

Data preprocessing is a crucial step in text-based machine learning tasks, particularly in sentiment analysis, as it enhances data quality and ensures that the input text is well-structured for algorithmic processing [32]. Raw textual data often contains inconsistencies, noise, and unnecessary elements that can negatively impact the performance of classification models. Therefore, systematic preprocessing is

DOI: IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

essential to improve text uniformity, reduce dimensionality, and optimize feature extraction. This study applies a series of preprocessing steps to refine the dataset before implementing machine learning algorithms. The following steps were conducted cleansing, casefolding, normalization, stopword removal and stemming. Each of these preprocessing techniques plays a vital role in refining the dataset for sentiment analysis.

2.4.1 *Cleansing:* is a fundamental preprocessing step in text mining that involves the removal of irrelevant or unnecessary elements from raw textual data. This process is essential in sentiment analysis, as unstructured data often contains noise that can negatively affect model performance [33]. By eliminating such noise, cleansing enhances text consistency and ensures that only meaningful information is retained for further analysis. In this research the cleansing process involved the removal of several unwanted components including user mentions such as "@ShopeeXpress" and "@ShopeeID", hashtags like "#ShopeeXpress", hyperlinks, emojis, special characters and Numerical values were removed unless they carried sentiment-related significance. Table 3 presents an example of the dataset before and after the cleansing process.

Table 3. Sample of Data After the Cleansing Process

Before	After
@ShopeeXpres @ShopeeID Gokil sih pakai ekspedisi spx 📦✨ paket nyampe cpt dari estimasi, kurirnya juga ramah. rekomendasi banget dan bakal pakai terus buat blnja online di shopee! 👍 #ShopeeXpress #Rekomendasi	Gokil sih pakai ekspedisi spx paket nyampe cpt dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat blnja online di shopee
@ShopeeCare @ShopeeID Pakai ShopeeXpress buat pertama kali, lumayan sih. Kadang cepat, kadang telat pket sampenya. Kurirnya baik, tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki.	Pakai ShopeeXpress buat pertama kali lumayan sih Kadang cepat kadang telat pket sampenya Kurirnya baik tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki
Tlng dong kurirnya lebih hati-hati! Barang yang aku pesan sampe pecah karena dilempar atau gimana gitu, sama ada bekas robekannya juga di kemasan paketnya. Jujur kaya kecewa banget sama plyanan dr ShopeeXpress, kalua gitu kayanya mending pake ekspedisi lain yg lebih aman.	Tlng dong kurirnya lebih hati-hati Barang yang aku pesan sampe pecah karena dilempar atau gimana gitu sama ada bekas robekannya juga di kemasan paketnya Jujur kaya kecewa banget sama plyanan dr ShopeeXpress kalua gitu kayanya mending pake ekspedisi lain yg lebih aman
Pernah coba buat retur barang pas pakai jasa ekspedissi ShopeeXpress. Prosesnya kagak yang terlalu ribet bangt sih, tapi lumayan lama juga nunggunya. Paketnya baru diambil kurir setelah 2 hari, terus butuh 5 hari lagi buat sampai ke seller. Jadi kalau butuh retur cepat bisa cari opsi lain.	Pernah coba buat retur barang pas pakai jasa ekspedissi ShopeeXpress Prosesnya kagak yang terlalu ribet bangt sih tapi lumayan lama juga nunggunya Paketnya baru diambil kurir setelah hari terus butuh hari lagi buat sampai ke seller Jadi kalau butuh retur cepat bisa cari opsi lain
@ShopeeID Sumpah kecewa bgt! pake jasa ekspedisi spx Paket udah 4 hari stuck di gudang, trs status brang kagak update sama sekali di aplikasinya, trs mana CS susah dihubungi pula. Tolong dibantu donk @ShopeeCare, ini barang urgent! lambat banget pengirimannya ☹ #ShopeeXpress #Lambat	Sumpah kecewa bgt pake jasa ekspedisi spx Paket udah hari stuck di gudang trs status brang kagak update sama sekali di aplikasinya, trs mana CS susah dihubungi pula Tolong dibantu donk ini barang urgent lambat banget pengirimannya
Awalnya ragu pakai jasa ekspedisi ShopeeXpres, tapi ternyata lumayan oke banget sih menurut gw. Gw pesen paket dari Jakarta ke Bandung cuma 5 hari dah sampe cuy, update tracking juga jelas di aplkasi Shopeenya. apalagi ada promo gratis ongkirnya dan paket sampe dengan aman. Semoga nggak turun kualitas pengirimannya ke depannya ya shopee! kasih bintang 5 dulu deh hehehehe ☆☆☆☆☆	Awalnya ragu pakai jasa ekspedisi ShopeeXpres tapi ternyata lumayan oke banget sih menurut gw Gw pesen paket dari Jakarta ke Bandung cuma hari dah sampe cuy update tracking juga jelas di aplkasi Shopeenya apalagi ada promo gratis ongkirnya dan paket sampe dengan aman Semoga nggak turun kualitas pengirimannya ke depannya ya shopee kasih bintang dulu deh hehehehe

2.4.2 *Casefolding:* is a text preprocessing technique that involves converting all characters in a text to lowercase [34]. This process standardizes textual data by eliminating variations in letter casing, ensuring that words with different capitalizations (e.g., "ShopeeXpress" and "shopeexpress") are treated as identical. The primary objective of casefolding is to minimize lexical discrepancies that may arise due to inconsistencies in capitalization. Without this step, machine learning models may incorrectly treat words with different letter cases as distinct entities, potentially leading to suboptimal feature extraction and model performance. Table 4 provides examples of text before and after the casefolding process.

Table 4. Sample of Data After the Casefolding Process

Before	After
Gokil sih pakai ekspedisi spx paket nyampe cpt dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat blnja online di shopee	gokil sih pakai ekspedisi spx paket nyampe cpt dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat blnja online di shopee
Pakai ShopeeXpress buat pertama kali lumayan sih Kadang cepat kadang telat pket sampenya Kurirnya baik tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki	pakai shopeexpress buat pertama kali lumayan sih kadang cepat kadang telat pket sampenya kurirnya baik tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki
Tlng dong kurirnya lebih hati-hati Barang yang aku pesan sampe pecah karena dilempar atau gimana gitu sama ada bekas robekannya juga di kemasan paketnya Jujur kaya kecewa banget sama plyanan dr ShopeeXpress kalua gitu kayanya mending pake ekspedisi lain yg lebih aman	tlng dong kurirnya lebih hati-hati barang yang aku pesan sampe pecah karena dilempar atau gimana gitu sama ada bekas robekannya juga di kemasan paketnya jujur kaya kecewa banget sama plyanan dr shopeexpress kalua gitu kayanya mending pake ekspedisi lain yg lebih aman
Pernah coba buat retur barang pas pakai jasa ekspedissi ShopeeXpress Prosesnya kagak yang terlalu ribet bangt sih tapi lumayan lama juga nunggunya Paketnya baru diambil kurir setelah hari terus butuh hari lagi buat sampai ke seller Jadi kalau butuh retur cepat bisa cari opsi lain	pernah coba buat retur barang pas pakai jasa ekspedissi shopeexpress prosesnya kagak yang terlalu ribet bangt sih tapi lumayan lama juga nunggunya paketnya baru diambil kurir setelah hari terus butuh hari lagi buat sampai ke seller jadi kalau butuh retur cepat bisa cari opsi lain
Sumpah kecewa bgt pake jasa ekspedisi spx Paket udah hari stuck di gudang trs status brang kagak	sumpah kecewa bgt pake jasa ekspedisi spx paket udah hari stuck di gudang trs status brang kagak

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

Before	After
update sama sekali di aplikasinya, trs mana CS susah dihubungi pula Tolong dibantu donk ini barang urgent lambat banget pengirimannya	update sama sekali di aplikasinya, trs mana cs susah dihubungi pula tolong dibantu donk ini barang urgent lambat banget pengirimannya
Awalnya ragu pakai jasa ekspedisi ShopeeXpres tapi ternyata lumayan oke banget sih menurut gw Gw pesen paket dari Jakarta ke Bandung cuma hari dah sampe cuy update tracking juga jelas di aplkasi Shopeenya apalagi ada promo gratis ongkirnya dan paket sampe dengan aman Semoga nggak turun kualitas pengirimannya ke depannya ya shopee kasih bintang dulu deh hehehehe	awalnya ragu pakai jasa ekspedisi shopeexpres tapi ternyata lumayan oke banget sih menurut gw gw pesen paket dari jakarta ke bandung cuma hari dah sampe cuy update tracking juga jelas di aplkasi shopeenya apalagi ada promo gratis ongkirnya dan paket sampe dengan aman semoga nggak turun kualitas pengirimannya ke depannya ya shopee kasih bintang dulu deh hehehehe

2.4.3 *Normalization:* is a text preprocessing technique aimed at standardizing words by converting non-standard or informal terms into their proper forms [35]. Normalization involves several transformations, including correction of typographical errors such as "cpt" are converted to "cepat", standardization of informal terms such as "gak" and "nggak" are mapped to their standard form "tidak", then unification of synonymous terms such as "ShopeeXpress" and "SPX" are standardized to a single representation to prevent inconsistencies in analysis. By ensuring that different variations of the same word are unified, the model can focus on analyzing sentiment patterns rather than being influenced by lexical inconsistencies. Table 5 illustrates examples of text before and after the normalization process.

Table 5. Sample of Data After the Normalization Process

Before	After
gokil sih pakai ekspedisi spx paket nyampe cpt dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat blnja online di shopee	gokil sih pakai ekspedisi spx paket sampai cepat dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat belanja online di shopee
pakai shopeexpress buat pertama kali lumayan sih kadang cepat kadang telat pket sampenya kurirnya baik tapi sistem tracking kadang juga nggak sesuai tolong di perbaiki	pakai spx buat pertama kali lumayan sih kadang cepat kadang telat paket sampainya kurirnya baik tapi sistem tracking kadang juga tidak sesuai tolong diperbaiki
tlng dong kurirnya lebih hati-hati barang yang aku pesan sampe pecah karena dilempar atau gimana gitu sama ada bekas robekannya juga di kemasan paketnya jujur kaya kecewa banget sama plyanan dr shopeexpress kalua gitu kayanya mending pake ekspedisi lain yg lebih aman	tolong dong kurirnya lebih hati-hati barang yang aku pesan sampai pecah entah karena dilempar atau bagaimana ditambah ada bekas robekan juga di kemasan paketnya jujur agak kecewa banget sama pelayanan dari spx kalau gitu kayanya mending pakai ekspedisi lain yang lebih aman
pernah coba buat retur barang pas pakai jasa ekspedissi shopeexpress prosesnya kagak yang terlalu ribet bangt sih tapi lumayan lama juga nunggunya paketnya baru diambil kurir setelah hari terus butuh hari lagi buat sampai ke seller jadi kalau butuh retur cepat bisa cari opsi lain	pernah coba buat retur barang pas pakai jasa ekspedisi spx prosesnya tidak terlalu ribet banget sih tapi lumayan lama juga nunggunya paketnya baru diambil kurir setelah beberapa hari terus butuh waktu lagi buat sampai ke seller jadi kalau butuh retur cepat bisa cari opsi lain
sumpah kecewa bgt pake jasa	sumpah kecewa banget pakai jasa

Before	After
ekspedisi spx paket udah hari stuck di gudang trs status brang kagak update sama sekali di aplikasinya, trs mana cs susah dihubungi pula tolong dibantu donk ini barang urgent lambat banget pengirimannya	ekspedisi spx paket sudah beberapa hari stuck di gudang terus status barang tidak update sama sekali di aplikasinya terus customer service susah dihubungi pula tolong dibantu dong ini barang urgent lambat banget pengirimannya
awalnya ragu pakai jasa ekspedisi shopeexpres tapi ternyata lumayan oke banget sih menurut gw gw pesen paket dari jakarta ke bandung cuma hari dah sampe cuy update tracking juga jelas di aplkasi shopeenya apalagi ada promo gratis ongkirnya dan paket sampe dengan aman semoga nggak turun kualitas pengirimannya ke depannya ya shopee kasih bintang dulu deh hehehehe	awalnya ragu pakai jasa ekspedisi spx tapi ternyata lumayan oke banget sih menurut gue gue pesan paket dari jakarta ke bandung cuma beberapa hari sudah sampai update tracking juga jelas di aplikasi shopee apalagi ada promo gratis ongkirnya dan paket sampai dengan aman semoga kualitas pengirimannya tidak turun ke depannya ya shopee kasih bintang dulu deh hehehehe

2.4.4 *Tokenizing:* is the process of breaking a text sequence into smaller units called tokens, which can be words, phrases, or subwords [36]. Tokenization plays a vital role in natural language processing (NLP) tasks, it allows the model to analyze text at the word level, improving its ability to detect patterns and relationships within the data. In this reaserch, tokenization was applied to the normalized text, where each sentence was split into its constituent words based on whitespace and punctuation rules. Table 6 presents examples of data before and after the tokenization process.

Table 6. Sample of Data After the Tokenizing Process

Before	After
gokil sih pakai ekspedisi spx paket sampai cepat dari estimasi kurirnya juga ramah rekomendasi banget dan bakal pakai terus buat belanja online di shopee	['gokil', 'sih', 'pakai', 'ekspedisi', 'spx', 'paket', 'sampai', 'cepat', 'dari', 'estimasi', 'kurirnya', 'juga', 'ramah', 'rekomendasi', 'banget', 'dan', 'bakal', 'pakai', 'terus', 'buat', 'belanja', 'online', 'di', 'shopee']
pakai spx buat pertama kali lumayan sih kadang cepat kadang telat paket sampainya kurirnya baik tapi sistem tracking kadang juga tidak sesuai tolong diperbaiki	['pakai', 'spx', 'buat', 'pertama', 'kali', 'lumayan', 'sih', 'kadang', 'cepat', 'kadang', 'telat', 'paket', 'sampainya', 'kurirnya', 'baik', 'tapi', 'sistem', 'tracking', 'kadang', 'juga', 'tidak', 'sesuai', 'tolong', 'diperbaiki']
tolong dong kurirnya lebih hati-hati barang yang aku pesan sampai pecah entah karena dilempar atau bagaimana ditambah ada bekas robekan juga di kemasan paketnya jujur agak kecewa banget sama pelayanan dari spx kalau gitu kayanya mending pakai ekspedisi lain yang lebih aman	['tolong', 'dong', 'kurirnya', 'lebih', 'hati-hati', 'barang', 'yang', 'aku', 'pesan', 'sampai', 'pecah', 'entah', 'karena', 'dilempar', 'atau', 'bagaimana', 'ditambah', 'ada', 'bekas', 'robekan', 'juga', 'di', 'kemasan', 'paketnya', 'jujur', 'agak', 'kecewa', 'banget', 'sama', 'pelayanan', 'dari', 'spx', 'kalau', 'gitu', 'kayanya', 'mending', 'pakai', 'ekspedisi', 'lain', 'yang', 'lebih', 'aman']
pernah coba buat retur barang pas pakai jasa ekspedisi spx prosesnya tidak terlalu ribet banget sih tapi lumayan lama juga nunggunya paketnya baru diambil kurir setelah beberapa hari terus butuh waktu lagi buat sampai ke seller jadi kalau butuh retur cepat bisa cari opsi lain	['pernah', 'coba', 'buat', 'retur', 'barang', 'pas', 'pakai', 'jasa', 'ekspedisi', 'spx', 'prosesnya', 'tidak', 'terlalu', 'ribet', 'banget', 'sih', 'tapi', 'lumayan', 'lama', 'juga', 'nunggunya', 'paketnya', 'baru', 'diambil', 'kurir', 'setelah', 'beberapa', 'hari', 'terus', 'butuh', 'waktu', 'lagi', 'buat', 'sampai', 'ke', 'seller', 'jadi', 'kalau', 'butuh', 'retur', 'cepat', 'bisa', 'cari', 'opsi', 'lain']
sumpah kecewa banget pakai	['sumpah', 'kecewa', 'banget', 'pakai',

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

Before	After
jasa ekspedisi spx paket sudah beberapa hari stuck di gudang terus status barang tidak update sama sekali di aplikasinya terus customer service susah dihubungi pula tolong dibantu dong ini barang urgent lambat banget pengirimannya	'jasa', 'ekspedisi', 'spx', 'paket', 'sudah', 'beberapa', 'hari', 'stuck', 'di', 'gudang', 'terus', 'status', 'barang', 'tidak', 'update', 'sama', 'sekali', 'di', 'aplikasinya', 'terus', 'customer', 'service', 'susah', 'dihubungi', 'pula', 'tolong', 'dibantu', 'dong', 'ini', 'barang', 'urgent', 'lambat', 'banget', 'pengirimannya']
awalnya ragu pakai jasa ekspedisi spx tapi ternyata lumayan oke banget sih menurut gue gue pesan paket dari jakarta ke bandung cuma beberapa hari sudah sampai update tracking juga jelas di aplikasi shopee apalagi ada promo gratis ongkirnya dan paket sampai dengan aman semoga kualitas pengirimannya tidak turun ke depannya ya shopee kasih bintang dulu deh hehehehe	['awalnya', 'ragu', 'pakai', 'jasa', 'ekspedisi', 'spx', 'tapi', 'ternyata', 'lumayan', 'oke', 'banget', 'sih', 'menurut', 'gue', 'gue', 'pesan', 'paket', 'dari', 'jakarta', 'ke', 'bandung', 'cuma', 'beberapa', 'hari', 'sudah', 'sampai', 'update', 'tracking', 'juga', 'jelas', 'di', 'aplikasi', 'shopee', 'apalagi', 'ada', 'promo', 'gratis', 'ongkirnya', 'dan', 'paket', 'sampai', 'dengan', 'aman', 'semoga', 'kualitas', 'pengirimannya', 'tidak', 'turun', 'ke', 'depannya', 'ya', 'shopee', 'kasih', 'bintang', 'dulu', 'deh', 'hehehehe']

2.4.5 *Stopword Removal:* is the process of eliminating common words that do not contribute significantly to the sentiment or meaning of a text [37]. These words, known as stopwords, typically include conjunctions, prepositions, pronouns, and other high-frequency words such as "dan", "di", "aku", and "juga". By eliminating these words, sentiment analysis models can concentrate on the most relevant features, thereby enhancing the accuracy of classification. In this research, stopword removal was performed using the Natural Language Toolkit (NLTK) and the Sastrawi library in Python. The stopword list used was based on predefined Indonesian stopwords, supplemented by custom stopwords relevant to this research. Table 7 provides examples of data before and after the stopword removal process.

Table 7. Sample of Data After the Stopword Removal Process

Before	After
['gokil', 'sih', 'pakai', 'ekspedisi', 'spx', 'paket', 'sampai', 'cepat', 'dari', 'estimasi', 'kurirnya', 'juga', 'ramah', 'rekomendasi', 'banget', 'dan', 'bakal', 'pakai', 'terus', 'buat', 'belanja', 'online', 'di', 'shopee']	['gokil', 'pakai', 'ekspedisi', 'spx', 'paket', 'sampai', 'cepat', 'estimasi', 'kurir', 'ramah', 'rekomendasi', 'belanja', 'online', 'shopee']
['pakai', 'spx', 'buat', 'pertama', 'kali', 'lumayan', 'sih', 'kadang', 'cepat', 'kadang', 'telat', 'paket', 'sampainya', 'kurirnya', 'baik', 'tapi', 'sistem', 'tracking', 'kadang', 'juga', 'tidak', 'sesuai', 'tolong', 'diperbaiki']	['pakai', 'spx', 'pertama', 'kali', 'kadang', 'cepat', 'kadang', 'telat', 'paket', 'sampai', 'kurir', 'baik', 'sistem', 'tracking', 'kadang', 'sesuai', 'tolong', 'diperbaiki']
['tolong', 'dong', 'kurirnya', 'lebih', 'hati-hati', 'barang', 'yang', 'aku', 'pesan', 'sampai', 'pecah', 'entah', 'karena', 'dilempar', 'atau', 'bagaimana', 'ditambah', 'ada', 'bekas', 'robekan', 'juga', 'di', 'kemasan', 'paketnya', 'jujur', 'agak', 'kecewa', 'banget', 'sama', 'pelayanan', 'dari', 'spx', 'kalau', 'gitu', 'kayanya', 'mending', 'pakai', 'ekspedisi', 'lain', 'yang', 'lebih', 'aman']	['tolong', 'kurir', 'hati-hati', 'barang', 'pesan', 'sampai', 'pecah', 'dilempar', 'bekas', 'robekan', 'kemasan', 'paket', 'kecewa', 'pelayanan', 'spx', 'pakai', 'ekspedisi', 'aman']
['pernah', 'coba', 'buat', 'retur', 'barang', 'pas', 'pakai', 'jasa', 'ekspedisi', 'spx', 'prosesnya', 'tidak', 'terlalu', 'ribet', 'banget', 'sih', 'tapi', 'lumayan', 'lama', 'juga', 'nunggunya', 'paketnya', 'baru', 'diambil', 'kurir', 'setelah', 'beberapa', 'hari', 'terus', 'butuh', 'waktu', 'lagi', 'buat', 'sampai', 'ke', 'seller', 'jadi', 'kalau', 'butuh', 'retur', 'cepat', 'bisa', 'cari', 'opsi', 'lain']	['coba', 'retur', 'barang', 'pakai', 'jasa', 'ekspedisi', 'spx', 'proses', 'ribet', 'lama', 'nunggu', 'paket', 'diambil', 'kurir', 'hari', 'butuh', 'waktu', 'sampai', 'seller', 'retur', 'cepat', 'cari', 'opsi']
['sumpah', 'kecewa', 'banget', 'pakai', 'jasa', 'ekspedisi', 'spx', 'paket', 'sudah', 'beberapa', 'hari', 'stuck', 'di', 'gudang', 'terus', 'status', 'barang', 'tidak', 'update', 'sama', 'sekali', 'di', 'aplikasinya', 'terus', 'customer', 'service', 'susah', 'dihubungi', 'pula', 'tolong', 'dibantu', 'dong', 'ini', 'barang', 'urgent', 'lambat', 'banget', 'pengirimannya']	['kecewa', 'pakai', 'jasa', 'ekspedisi', 'spx', 'paket', 'hari', 'stuck', 'gudang', 'status', 'barang', 'update', 'aplikasi', 'customer', 'service', 'susah', 'dihubungi', 'tolong', 'dibantu', 'urgent', 'lambat', 'pengiriman']
['awalnya', 'ragu', 'pakai', 'jasa', 'ekspedisi', 'spx', 'tapi', 'ternyata', 'lumayan', 'oke', 'banget', 'sih', 'menurut', 'gue', 'gue', 'pesan', 'paket', 'dari', 'jakarta', 'ke', 'bandung', 'cuma', 'beberapa', 'hari', 'sudah', 'sampai', 'update', 'tracking', 'juga', 'jelas', 'di', 'aplikasi', 'shopee', 'apalagi', 'ada', 'promo', 'gratis', 'ongkirnya', 'dan', 'paket', 'sampai', 'dengan', 'aman', 'semoga', 'kualitas', 'pengirimannya', 'tidak', 'turun', 'ke', 'depannya', 'ya', 'shopee', 'kasih', 'bintang', 'dulu', 'deh', 'hehehehe']	['ragu', 'pakai', 'jasa', 'ekspedisi', 'spx', 'ternyata', 'oke', 'pesan', 'paket', 'jakarta', 'bandung', 'hari', 'sampai', 'update', 'tracking', 'jelas', 'aplikasi', 'shopee', 'promo', 'gratis', 'ongkir', 'paket', 'aman', 'kualitas', 'pengiriman', 'turun', 'shopee', 'bintang']

2.4.6 *Stemming:* is the process of reducing words to their root or base form by removing affixes such as prefixes, suffixes, infixes, or confixes [38]. This technique is essential in natural language processing (NLP) to standardize words with similar meanings, ensuring that variations of a word are treated as a single entity. For instance, words like "dilempar" (thrown), "pelayanan" (service), and "pemesanan" (ordering) would be transformed into their base forms: "lempar," "layan," and "pesan," respectively. In this research, stemming was conducted using the Sastrawi library in Python, which is specifically designed for stemming the Indonesian language. Sastrawi implements the Nazief-Adriani algorithm, a widely used stemming algorithm for Bahasa Indonesia. Table 8 presents examples of text data before and after the stemming process.

Table 8. Sample of Data After the Stemming Process

Before	After
['gokil', 'pakai', 'ekspedisi', 'spx', 'paket', 'sampai', 'cepat', 'estimasi', 'kurir', 'ramah', 'rekomendasi', 'belanja', 'online', 'shopee']	['gokil', 'pakai', 'ekspedisi', 'spx', 'paket', 'sampai', 'cepat', 'estimasi', 'kurir', 'ramah', 'rekomendasi', 'belanja', 'online', 'shopee']
['pakai', 'spx', 'pertama', 'kali', 'kadang', 'cepat', 'kadang', 'telat', 'paket', 'sampai', 'kurir', 'baik', 'sistem', 'tracking', 'kadang', 'sesuai', 'tolong', 'diperbaiki']	['pakai', 'spx', 'pertama', 'kali', 'kadang', 'cepat', 'kadang', 'telat', 'paket', 'sampai', 'kurir', 'baik', 'sistem', 'tracking', 'kadang', 'sesuai', 'tolong', 'perbaiki']
['tolong', 'kurir', 'hati-hati', 'barang', 'pesan', 'sampai', 'pecah', 'dilempar', 'bekas', 'robekan', 'kemasan', 'paket', 'kecewa', 'pelayanan', 'spx', 'pakai', 'ekspedisi', 'aman']	['tolong', 'kurir', 'hati-hati', 'barang', 'pesan', 'sampai', 'pecah', 'lempar', 'bekas', 'robek', 'kemas', 'paket', 'kecewa', 'layan', 'spx', 'pakai', 'ekspedisi', 'aman']
['coba', 'retur', 'barang', 'pakai', 'jasa', 'ekspedisi', 'spx', 'proses', 'ribet', 'lama', 'nunggu', 'paket', 'diambil', 'kurir', 'hari', 'butuh', 'waktu',	['coba', 'retur', 'barang', 'pakai', 'jasa', 'ekspedisi', 'spx', 'proses', 'ribet', 'lama', 'nunggu', 'paket', 'ambil', 'kurir', 'hari', 'butuh',

DOI: IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

Before	After
'sampai', 'seller', 'retur', 'cepat', 'cari', 'opsi']	'waktu', 'sampai', 'seller', 'retur', 'cepat', 'cari', 'opsi']
['kecewa', 'pakai', 'jasa', 'ekspedisi', 'spx', 'paket', 'hari', 'stuck', 'gudang', 'status', 'barang', 'update', 'aplikasi', 'customer', 'service', 'susah', 'dihubungi', 'tolong', 'dibantu', 'urgent', 'lambat', 'pengiriman']	['kecewa', 'pakai', 'jasa', 'ekspedisi', 'spx', 'paket', 'hari', 'stuck', 'gudang', 'status', 'barang', 'update', 'aplikasi', 'customer', 'service', 'susah', 'hubung', 'tolong', 'bantu', 'urgent', 'lambat', 'kirim']
['ragu', 'pakai', 'jasa', 'ekspedisi', 'spx', 'ternyata', 'oke', 'pesan', 'paket', 'jakarta', 'bandung', 'hari', 'sampai', 'update', 'tracking', 'jelas', 'aplikasi', 'shopee', 'promo', 'gratis', 'ongkir', 'paket', 'aman', 'kualitas', 'pengiriman', 'turun', 'shopee', 'bintang']	['ragu', 'pakai', 'jasa', 'ekspedisi', 'spx', 'ternyata', 'oke', 'pesan', 'paket', 'jakarta', 'bandung', 'hari', 'sampai', 'update', 'tracking', 'jelas', 'aplikasi', 'shopee', 'promo', 'gratis', 'ongkir', 'paket', 'aman', 'kualitas', 'kirim', 'turun', 'shopee', 'bintang']

2.5 *Algorithm Implementation*

After the preprocessing stage, the next step in this study involves feature extraction and the implementation of machine learning algorithms. In this research, the Term Frequency-Inverse Document Frequency (TF-IDF) method is employed to convert textual data into numerical representations that can be processed by machine learning models [39]. The TF-IDF formula shown in Equation 1 [40] is employed to assess the cumulative weight of a word in a document [41].

$$TF - IDF_{(t,d,D)} = TF_{(t,d)} \times IDF_{(t,D)} \quad (1)$$

The training data is used to build the model, while the testing data is utilized to evaluate its performance. Additionally, two classification algorithms, Support Vector Machine (SVM) and Logistic Regression (LR), are applied to classify sentiment into three categories: positive, neutral, and negative.

2.6 *Model Evaluation*

Evaluating the performance of a classification model is a crucial step in sentiment analysis to ensure its effectiveness in accurately categorizing textual data [42]. A well-defined evaluation process provides insights into the model's strengths and weaknesses, guiding improvements for better predictive performance. In this study, the confusion matrix is utilized as the foundation for performance evaluation. A confusion matrix is a tabular representation that compares the predicted classifications against the actual labels [43]. From the confusion matrix several performance metrics can be derived such as accuracy, recall, precision, and F1-score. Each metric provides valuable insight into the model's classification ability.

- *Accuracy:* measures the proportion of correctly classified instances to the total number of instances [44]. It is defined at Equation 2 [45]:

$$Accuracy = \frac{(TP_{positive} + TP_{negative} + TP_{neutral})}{(Total\ Cases)} \quad (2)$$

- *Recall:* quantifies the model's ability to correctly identify all instances of a given sentiment [46]. It is calculated using Equation 3 [47]:

$$Recall = \frac{(TP)}{(TP+FN)} \quad (3)$$

- *Precision:* measures the proportion of correctly classified positive instances out of all instances predicted as positive [48]. It is defined at Equation 4 [49]:

$$Precision = \frac{(TP)}{(TP+FP)} \quad (4)$$

- *The F1-score:* is the harmonic mean of precision and recall, providing a balanced evaluation of both metrics [50]. It is expressed at Equation 5 [51]:

$$F1\ score = 2\ x \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (5)$$

By employing these evaluation techniques, this study ensures that the sentiment classification model is assessed comprehensively, leading to an informed selection of the most effective algorithm.

3 RESULT AND DISCUSSION

This section presents the findings of the sentiment analysis on customer reviews of Shopee Xpress, along with an in-depth discussion of the results. The analysis covers sentiment distribution, word frequency patterns, model performance evaluation, and a comparative assessment of classification models. Furthermore, a quantitative evaluation of the models is conducted to ensure the reliability of the sentiment classification. The insights derived from this study provide valuable implications for improving Shopee Xpress services and understanding customer satisfaction trends.

3.1 *Overview of Sentiment Analysis Results*

This section provides a general overview of the sentiment classification results obtained from the dataset, which consists of 2480 customer reviews on Shopee Xpress. Fig. 2 illustrates the distribution of sentiment within the dataset, where the classification process categorizes sentiments into three classes: positive, neutral, and negative.

DOI: IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

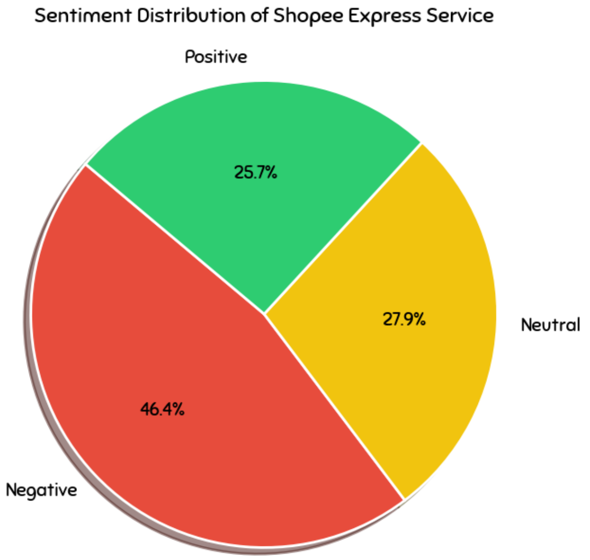

Figure 2. Sentiment distribution in the dataset

As depicted in Fig. 2, the majority of customer reviews exhibit a negative sentiment, accounting for 1151 instances (46.41%) of the total dataset. Meanwhile, neutral sentiment comprises 691 reviews (27.86%), and positive sentiment is the least represented, with 638 reviews (25.73%).

The predominance of negative sentiment in the dataset suggests that a substantial number of customers have expressed dissatisfaction with Shopee Xpress services. This finding may indicate recurring issues such as delays in delivery, lost packages, or poor handling of customer complaints. The significant proportion of neutral reviews implies that a considerable segment of users remains undecided or provides feedback without strong emotional polarity. Conversely, the lower proportion of positive sentiment reflects that only a quarter of the customers reported a satisfactory experience with the service.

One possible correlation between sentiment distribution and customer satisfaction is that a higher volume of negative sentiment may directly reflect operational inefficiencies or unmet customer expectations. The relatively high percentage of neutral sentiment suggests that many customers either express factual feedback without emotional inclination or experience mixed service quality. This could indicate an opportunity for Shopee Xpress to engage with neutral reviewers and convert them into satisfied customers by addressing their concerns proactively.

From a business perspective, Shopee Xpress can leverage these insights to enhance service quality. Addressing the major pain points highlighted in negative reviews and enhancing customer support may lead to a reduction in negative sentiment over time. Additionally, proactive customer engagement strategies, such as personalized assistance or compensation for service failures, could contribute to a higher proportion of positive feedback in the future. Furthermore, analyzing the specific aspects of service praised in positive reviews can help reinforce existing strengths and build stronger customer loyalty. Overall, these findings underscore the importance of continuous monitoring of customer feedback to drive service improvements.

3.2 *Sentiment Distribution Over Time*

Understanding how customer sentiment fluctuates over time is crucial in identifying patterns and potential external factors influencing user experiences. Fig. 3 presents a bar chart illustrating the monthly distribution of sentiment from January 2024 to March 2025, highlighting trends in positive, neutral, and negative sentiments.

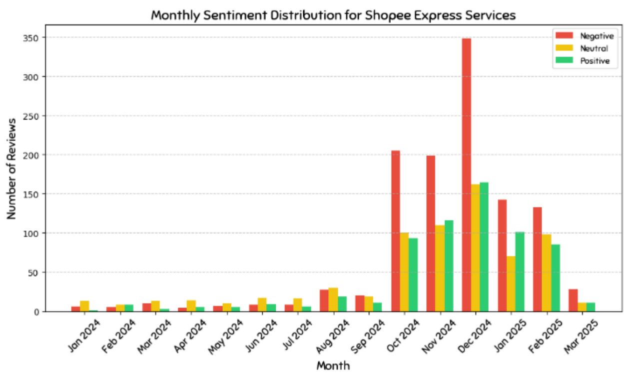

Figure 3. Sentiment distribution by monthly

The distribution of sentiment in Fig. 3 exhibits noticeable variations across different months. Early in the dataset (January–July 2024), sentiment fluctuations appear relatively stable, with negative, neutral, and positive reviews occurring in relatively low and balanced proportions. However, a significant increase in review volume is observed starting in August 2024, with a substantial rise in negative sentiment in October, November, and December 2024. Specifically, negative sentiment reached its peak in December 2024, with 349 negative reviews, marking the highest count recorded within the dataset. This trend suggests potential dissatisfaction during peak periods, which may coincide with high-order volumes due to promotional events or holiday shopping seasons.

Conversely, neutral and positive sentiments also increased proportionally during the same period, albeit at a lower rate than negative sentiment. The relatively high number of neutral reviews suggests that some customers experienced mixed service quality, while positive reviews indicate that a portion of users remained satisfied despite overall concerns.

Potential Insights and Business Implications, there are several key insights can be drawn from the sentiment distribution over time:

- *Correlation with External Factors:* The surge in negative sentiment in the last quarter of 2024 (October–December) may be associated with increased demand during major shopping events, such as year-end sales, where logistical challenges

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

and service delays commonly occur. Understanding these patterns can help Shopee Xpress anticipate potential service disruptions and improve operational efficiency during peak periods.

- *Improvement Opportunities:* By analyzing months with high negative sentiment, Shopee Xpress can investigate specific service issues that contributed to dissatisfaction. For example, common complaints in these periods may relate to delivery delays, damaged goods, or customer service inefficiencies. Addressing these concerns proactively can enhance customer satisfaction and retention.
- *Long-Term Trend Monitoring:* The decline in negative sentiment from January to March 2025 suggests potential improvements in service quality or a decrease in order volume post-holiday season. Continuously monitoring sentiment trends can provide valuable feedback for strategic decision-making and service enhancement.

By leveraging these insights, Shopee Xpress can implement targeted improvements to optimize service performance, especially during peak demand periods, and enhance the overall customer experience.

3.3 *Word Frequency Analysis in Customer Reviews*

Analyzing word frequency in customer reviews provides valuable insights into prevalent themes and recurring concerns expressed by users. By identifying the most frequently occurring words, it is possible to understand the primary topics discussed by customers, which can reflect their experiences, expectations, and pain points. This section presents a detailed analysis of word frequency across different sentiment categories, offering a deeper understanding of the language patterns in customer feedback related to Shopee Xpress.

3.3.1 *Overall Word Frequency:* Text analysis is a fundamental technique in sentiment analysis, providing insights into the most frequently occurring words within a dataset. One of the most effective visual representations of word frequency is a word cloud , which displays words in varying sizes based on their occurrence rates [52]. The larger the word appears in the visualization, the more frequently it occurs in customer reviews. Fig. 4 illustrates the word cloud generated from all collected reviews.

Figure 4. Word cloud of customer reviews on shopee xpress

The word cloud reveals in Fig. 4 the dominant words used by customers when expressing their experiences with Shopee Xpress. The most frequently occurring word is "spx" which is an abbreviation commonly used for Shopee Xpress. This indicates that customers explicitly mention the courier service in their reviews, reflecting direct engagement with the brand. Other frequently mentioned words include "kirim", "shopee" and "paket". These words highlight the central themes of customer discussions, primarily focusing on the delivery process and the e-commerce platform. Additionally, words like "kurir", "barang" and "pakai" suggest that many reviews discuss courier performance and package conditions. Notably, the word "lambat" appears frequently, indicating potential customer dissatisfaction with delivery speed. This finding suggests that delivery delays might be a recurring issue in customer experiences with Shopee Xpress.

Understanding word frequency distribution provides valuable insights into customer concerns and expectations. The prominence of words related to delivery, courier services, and delays suggests that customers primarily focus on service efficiency. Additionally, these findings can aid Shopee Xpress in refining their service by prioritizing aspects that are frequently discussed by users, such as courier reliability and shipment timeliness. By leveraging these insights, Shopee Xpress can enhance customer satisfaction by identifying key areas for improvement, such as optimizing delivery operations and improving communication regarding shipping times.

3.3.2 *Frequent Words in Negative Sentiment:* Understanding the key terms frequently appearing in negative sentiment reviews provides valuable insights into customer dissatisfaction and recurring service issues. Fig. 5 presents a bar chart visualizing the 25 most frequently occurring words in negative sentiment reviews, offering a clearer perspective on the dominant themes of customer complaints.

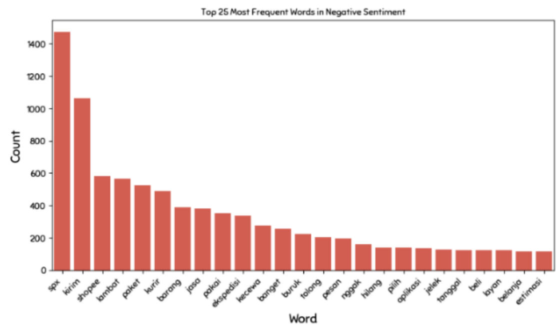

Figure 5. Frequency of words in negative sentiment

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

The most frequently mentioned word in negative reviews in Fig. 5 is "spx" which refers to Shopee Xpress. This indicates that many users explicitly mention the courier service when expressing dissatisfaction. The prominence of "kirim", "shopee", and "lambat" highlights that delivery speed is a major concern among customers. The presence of "paket" and "kurir" suggests that issues related to package handling and courier service efficiency are also frequently discussed. Several words further emphasize negative customer experiences, such as "kecewa", "buruk", and "hilang". These terms indicate dissatisfaction with service quality, lost or delayed shipments, and unmet expectations. Additionally, "tolong" frequently appears, suggesting that many customers request assistance, likely due to unresolved complaints or poor customer support. The frequent appearance of words like "tanggal" and "estimasi" suggests that delays and inaccurate estimated delivery times are common grievances. Similarly, "aplikasi" may indicate dissatisfaction with the Shopee platform’s logistics tracking system or customer service responsiveness. The word "pilih" may reflect customer frustration regarding the inability to select alternative shipping options or dissatisfaction with forced courier assignment.

By leveraging these insights, Shopee Xpress can develop targeted solutions to address common pain points, such as enhancing delivery efficiency, improving customer service responsiveness, and refining estimated delivery time accuracy.

3.3.3 *Frequent Words in Neutral Sentiment:* Neutral sentiment reviews provide insights into customer experiences that are neither highly positive nor strongly negative, often reflecting transactional or factual statements rather than explicit satisfaction or dissatisfaction. Fig. 6 presents a bar chart illustrating the 25 most frequently occurring words in neutral sentiment reviews, which helps in understanding the nature of these reviews and their implications for Shopee Xpress.

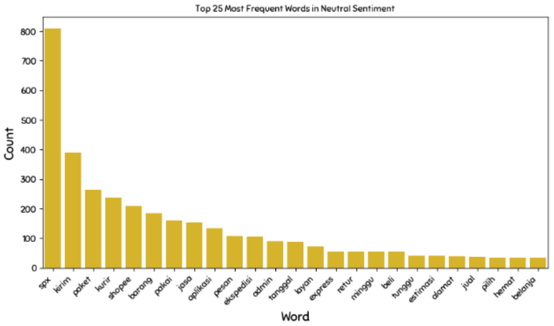

Figure 6. Frequency of words in neutral sentiment

The most frequently in Fig. 6 mentioned term is "spx" indicating that Shopee Xpress is commonly referenced in neutral reviews. Words such as "kirim", "paket" and "kurir" suggest that many neutral reviews focus on descriptive aspects of the shipping process rather than opinions on service quality. Interestingly, words like "aplikasi", "admin" and "layan" suggest that some neutral reviews may relate to the technical aspects of Shopee's platform or customer service interactions, where users might be describing their experiences without strong emotional expressions. Additionally, the presence of terms like "tanggal", "estimasi" and "tunggu" (wait, 42 occurrences) may indicate discussions about delivery timeframes, where customers acknowledge estimated delivery schedules but do not necessarily express frustration or praise. A significant portion of neutral reviews appears to contain informational or status updates rather than clear feedback. This may indicate that some customers are uncertain about their overall satisfaction with the service or simply reporting their experiences without forming a strong opinion. For instance, terms such as "retur" and "hemat" suggest that some users discuss order processing, returns, or cost-saving aspects of the service in a matter-of-fact manner, rather than framing them as positive or negative experiences.

The prevalence of neutral sentiment suggests that a portion of customers neither express strong dissatisfaction nor high satisfaction, which may indicate an opportunity for Shopee Xpress to engage with these customers further. Understanding the context of these reviews could help Shopee Xpress enhance service differentiation by proactively addressing concerns before they escalate into negative sentiment. Furthermore, given that many neutral reviews mention delivery times, package handling, and courier services, Shopee Xpress could improve communication transparency by ensuring that customers receive clearer updates regarding delivery progress. Encouraging more detailed feedback through follow-up surveys or personalized support could also help convert neutral experiences into positive ones, ultimately improving overall customer satisfaction.

3.3.4 *Frequent Words in Positive Sentiment:* Understanding the most frequently used words in positive sentiment reviews provides valuable insights into the key factors contributing to customer satisfaction with Shopee Xpress. Fig. 7 presents a bar chart illustrating the 25 most frequently occurring words in positive sentiment reviews, highlighting common themes associated with favorable customer experiences.

10

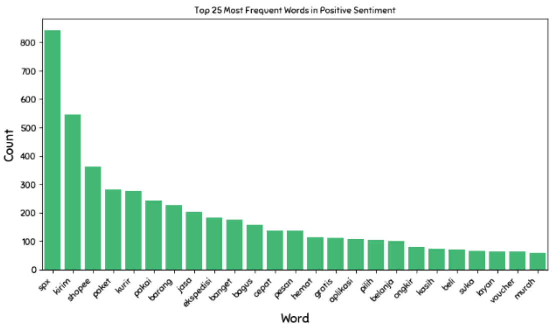

Figure 7. Frequency of words in positive sentiment

The most frequently mentioned word in Fig. 7 is "spx" indicating that Shopee Xpress is a central focus in positive reviews. Words such as "kirim", "shopee", "paket" and "kurir" suggest that fast and reliable delivery services are a significant factor contributing to positive feedback. Furthermore, terms like "cepat", "bagus" and "banget" reinforce that speed and service quality are highly valued by customers. Additionally, "gratis", "hemat" and "voucher" suggest that promotional offers, such as free shipping and discounts, play a role in enhancing customer satisfaction.

The analysis of positive sentiment words indicates that fast delivery, affordability, and good service quality are the primary drivers of customer satisfaction. Since terms like "cepat" and "bagus" frequently appear in positive reviews, Shopee Xpress should continue to prioritize efficient delivery operations. Improving logistics management and optimizing courier networks could further strengthen customer trust. The presence of words like "gratis", "hemat", and "voucher" highlights the importance of affordability and discounts in shaping positive customer experiences. Shopee Xpress could enhance engagement by offering more targeted promotions, such as loyalty-based free shipping or seasonal discount campaigns. Words like "layan" and "kasih" indicate that personalized customer service efforts are appreciated. Investing in responsive customer support and proactive communication may further boost overall satisfaction. By aligning service improvements with these insights, Shopee Xpress can sustain and expand its positive reputation while addressing potential areas for further enhancement.

3.4 *Model Performance Evaluation*

Evaluating model performance is crucial in determining the effectiveness of sentiment classification. This section examines the classification capabilities of the models by analyzing their confusion matrices, which provide insight into the distribution of correctly and incorrectly classified sentiment labels. The analysis offers a deeper understanding of how well each model distinguishes between negative, neutral, and positive sentiments.

3.4.1 *Confusion Matrix Analysis for SVM:* The confusion matrix provides a detailed evaluation of the Support Vector Machine (SVM) model's classification performance in predicting sentiment categories. Fig. 8 illustrates the confusion matrix, summarizing the number of correctly and incorrectly classified sentiment labels across three categories: negative, neutral, and positive.

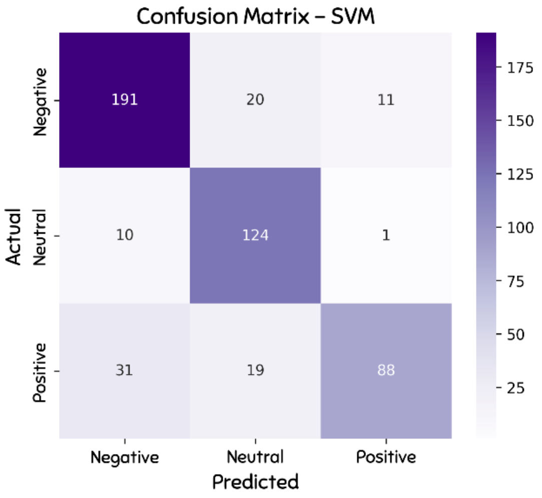

Figure 8. Confusion matrix of svm algorithm

Each cell in the confusion matrix at Fig. 8 represents a specific classification outcome:

- *Negative Sentiment Classification:* The model correctly identified 191 negative reviews (True Negative). However, 20 negative reviews were misclassified as neutral, and 11 were misclassified as positive. This suggests that the model is generally effective in detecting negative sentiment, but there is some overlap with neutral sentiment, possibly due to ambiguous expressions or mixed emotions.
- *Neutral Sentiment Classification:* 124 reviews were correctly classified as neutral, showing strong performance in recognizing neutrality. However, 10 neutral reviews were misclassified as negative, and 1 was misclassified as positive. The low misclassification rate indicates that the model performs well in identifying neutral sentiment, likely because these reviews contain balanced expressions without strong polarity.
- *Positive Sentiment Classification:* The model correctly classified 88 positive reviews (True Positive). However, 31 positive reviews were misclassified as

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

negative, and 19 were misclassified as neutral. The relatively high misclassification rate for positive sentiment suggests that the model struggles to differentiate between positive and other sentiment categories, possibly due to variations in expression intensity or the presence of minor complaints in otherwise positive reviews.

3.4.2 *Confusion Matrix Analysis for Logistic Regression:* The confusion matrix provides an in-depth assessment of the Logistic Regression model's classification performance in predicting sentiment categories. Fig. 9 presents the confusion matrix, summarizing the number of correctly and incorrectly classified sentiment labels across three categories: negative, neutral, and positive.

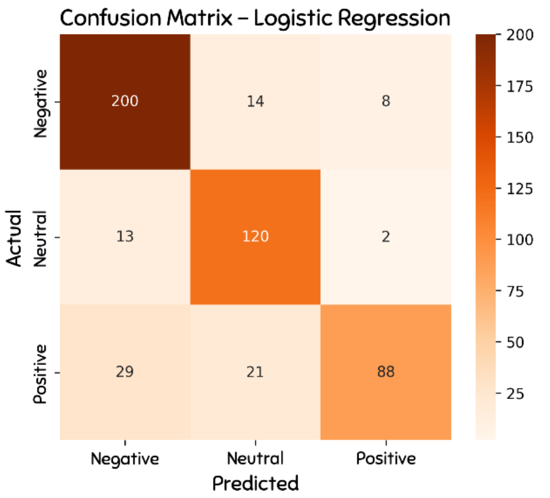

Figure 9. Confusion matrix of logistic regression algorithm

Each cell in the confusion matrix at Fig.9 represents a specific classification outcome:

- *Negative Sentiment Classification:* The model correctly identified 200 negative reviews (True Negative). However, 14 negative reviews were misclassified as neutral, and 8 were misclassified as positive. The relatively low misclassification rate indicates that the model effectively distinguishes negative sentiments from other categories, although there is a minor overlap with neutral sentiment.

- *Neutral Sentiment Classification:* The model correctly classified 120 neutral reviews (True Neutral). However, 13 neutral reviews were misclassified as negative, and 2 were misclassified as positive. The misclassification of neutral reviews as negative suggests that some reviews might contain mildly negative phrases or mixed expressions, leading to incorrect classification.

- *Positive Sentiment Classification:* The model correctly classified 88 positive reviews (True Positive). However, 29 positive reviews were misclassified as negative, and 21 were misclassified as neutral. Similar to SVM, the misclassification rate for positive sentiment is relatively high, indicating that the model struggles to differentiate positive sentiment from neutral or negative expressions

3.5 *Model Comparison on Unlabeled Data*

To further evaluate the performance and consistency of the SVM and Logistic Regression models, both classifiers were applied to a set of 620 unlabeled reviews. The results of this classification are illustrated in Fig. 10, which presents a bar chart comparing the sentiment distribution predicted by each model.

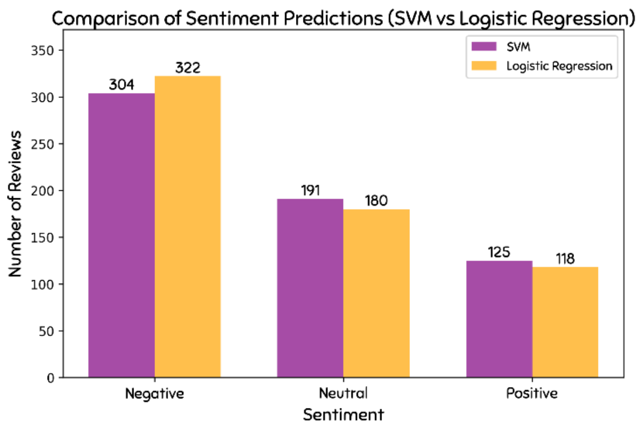

Figure 10. Sentiment distribution across models

While both models in Fig. 10 exhibit similar classification patterns, notable differences can be observed in the distribution of sentiment predictions, particularly in the negative and neutral categories. The classification results for SVM and Logistic Regression on the unlabeled dataset are summarized as follows:

- *Negative Sentiment Predictions:* Logistic Regression classified 322 reviews as negative, while SVM classified 304 reviews as negative indicating a slightly higher tendency of Logistic Regression to assign negative sentiment. This difference suggests that Logistic Regression may have a lower threshold for identifying negative sentiment, potentially misclassifying neutral reviews as negative.

- *Neutral Sentiment Predictions:* SVM assigned 191 reviews to the neutral category, whereas Logistic Regression classified 180 as neutral. The slightly lower neutral classification in Logistic Regression

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

could indicate that some neutral reviews were classified as either negative or positive, reflecting a more polarized classification tendency in the Logistic Regression model.

- *Positive Sentiment Predictions:* Both models classified the least number of reviews as positive, with SVM assigning 125 and Logistic Regression assigning 118 to this category. The minor difference suggests that both models face similar challenges in distinguishing positive sentiment, possibly due to lexical similarities between neutral and positive reviews in the dataset.

3.6 *Quantitative Evaluation of Models*

60 A comprehensive quantitative evaluation is essential to objectively assess the performance of sentiment
69 classification models. This section presents a comparative analysis of key classification metrics and examines the consistency between manually calculated and programmatically computed evaluation scores. The analysis
47 aims to ensure the accuracy and reliability of the models in classifying sentiment within the given dataset.

12 3.6.1 *Comparison of Classification Metrics:* To evaluate the performance of both classification models (SVM and Logistic Regression), a comparison of
42 key classification metrics like True Positive (TP), False Positive (FP), and False Negative (FN) was conducted. These values are summarized in Table 9, derived from the confusion matrices of each model.

Table 9. Comparison of TP, FP, and FN Values for Each Model

Sentiment	Metric	SVM	Logistic Regression
Negative	TP (True Positive)	191	200
	FP (False Positive)	41	42
	FN (False Negative)	31	22
Neutral	TP (True Positive)	124	120
	FP (False Positive)	39	35
	FN (False Negative)	11	15
Positive	TP (True Positive)	88	88
	FP (False Positive)	12	10
	FN (False Negative)	50	50

3.6.2 *Manual vs. Programmatic Calculation of Metrics:* To ensure the reliability of the evaluation metrics, this section compares the manual calculations of classification metrics with programmatically computed values using Python's sklearn.metrics
19 library. Referring to the data presented in Table 9, a
1 manual calculation is performed to obtain the accuracy, recall, precision, and F1-score for both
49 models. This calculation allows for a more precise evaluation of each model's performance in sentiment classification.

- *Accuracy:*

$$Accuracy = \frac{(TP_{positive} + TP_{negative} + TP_{neutral})}{(Total\ Cases)}$$

$$SVM = \frac{88 + 191 + 124}{495} = 0.814\ or\ 81\%$$

$$LR = \frac{88 + 200 + 120}{495} = 0.824\ or\ 82\%$$

- Recall Negative:

$$Recall_{negative} = \frac{(TP_{negative})}{(TP_{negative} + FN_{negative})}$$

$$SVM = \frac{191}{191 + 31} = 0.860\ or\ 86\%$$

$$LR = \frac{200}{200 + 22} = 0.900\ or\ 90\%$$

- Recall Neutral:

$$Recall_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral} + FN_{neutral})}$$

$$SVM = \frac{124}{124 + 11} = 0.918\ or\ 92\%$$

$$LR = \frac{120}{120 + 15} = 0.888\ or\ 89\%$$

- Recall Positive:

$$Recall_{positive} = \frac{(TP_{positive})}{(TP_{positive} + FN_{positive})}$$

$$SVM = \frac{88}{88 + 50} = 0.637\ or\ 64\%$$

$$LR = \frac{88}{88 + 50} = 0.637\ or\ 64\%$$

- Precision Negative:

$$Precision_{negative} = \frac{(TP_{negative})}{(TP_{negative} + FP_{negative})}$$

$$SVM = \frac{191}{191 + 41} = 0.823\ or\ 82\%$$

$$LR = \frac{200}{200 + 42} = 0.826\ or\ 83\%$$

- Precision Neutral:

DOI:
IJID (International Journal on Informatics for Development), e-ISSN: 2549-7448
Vol. 12, No. 1, 2023, Pp. X-Y

$$Precision_{neutral} = \frac{(TP_{neutral})}{(TP_{neutral} + FP_{neutral})}$$

$$SVM = \frac{124}{124 + 39} = 0.760 \ or \ 76\%$$

$$LR = \frac{120}{120 + 35} = 0.774 \ or \ 77\%$$

- Precision Positive:

$$Precision_{positive} = \frac{(TP_{positive})}{(TP_{positive} + FP_{positive})}$$

$$SVM = \frac{88}{88 + 12} = 0.880 \ or \ 88\%$$

$$LR = \frac{88}{88 + 10} = 0.897 \ or \ 90\%$$

- F1-score Negative:

$$F1\ score_{negative} = 2\,x \frac{(Precision_{negative}\ x\ Recall_{negative})}{(Precision_{negative} + Recall_{negative})}$$

$$SVM = 2x \frac{(0.823\ x\ 0.860)}{(0.823 + 0.860)} = 0.841 \ or \ 84\%$$

$$LR = 2x \frac{(0.826\ x\ 0.90)}{(0.826 + 0.90)} = 0.862 \ or \ 86\%$$

- F1-score Neutral:

$$F1\ score_{neutral} = 2\,x \frac{(Precision_{neutral}\ x\ Recall_{neutral})}{(Precision_{neutral} + Recall_{neutral})}$$

$$SVM = 2x \frac{(0.760\ x\ 0.918)}{(0.760 + 0.918)} = 0.832 \ or \ 83\%$$

$$LR = 2x \frac{(0.774\ x\ 0.888)}{(0.774 + 0.888)} = 0.827 \ or \ 83\%$$

- F1-score Positive:

$$F1\ score_{positive} = 2\,x \frac{(Precision_{positive}\ x\ Recall_{positive})}{(Precision_{positive} + Recall_{positive})}$$

$$SVM = 2x \frac{(0.880\ x\ 0.637)}{(0.880 + 0.637)} = 0.739 \ or \ 74\%$$

$$LR = 2x \frac{(0.897\ x\ 0.637)}{(0.897 + 0.637)} = 0.745 \ or \ 75\%$$

Fig. 11 illustrates the calculated accuracy, recall, precision, and F1-score for the Support Vector Machine (SVM) and Logistic Regression algorithms. These metrics were obtained using Python, utilizing the sklearn.metrics library to ensure an accurate and consistent evaluation of model performance.

```
==== SVM ====
Accuracy: 0.8141
              precision    recall  f1-score   support

    Negative       0.82      0.86      0.84       222
     Neutral       0.76      0.92      0.83       135
    Positive       0.88      0.64      0.74       138

    accuracy                           0.81       495
   macro avg       0.82      0.81      0.80       495
weighted avg       0.82      0.81      0.81       495


==== Logistic Regression ====
Accuracy: 0.8242
              precision    recall  f1-score   support

    Negative       0.83      0.90      0.86       222
     Neutral       0.77      0.89      0.83       135
    Positive       0.90      0.64      0.75       138

    accuracy                           0.82       495
   macro avg       0.83      0.81      0.81       495
weighted avg       0.83      0.82      0.82       495
```

Figure 11. Performance metrics of svm and logistic regression models

A comparison of the results reveals that both manual and programmatic calculations, as illustrated in Fig. 11, produce identical values for all four metrics. This consistency validates the accuracy of the evaluation process, confirming that the implementation is precise and devoid of computational errors.

3.7 *Final Model Selection and Business Insights*

The final step in this study involves selecting the most suitable sentiment classification model based on key performance metrics. Table 10 presents a comparative analysis of accuracy, recall, precision, and F1-score for both SVM and Logistic Regression.

Table 10. Comparison of SVM and Logistic Regression Model Evaluations

Metric	SVM	Logistic Regression
Accuracy	81	82
Recall (Negative)	86	90
Recall (Neutral)	92	89
Recall (Positive)	64	64
Precision (Negative)	82	83
Precision (Neutral)	76	77

Metric	SVM	Logistic Regression
Precision (Positive)	88	90
F1-score (Negative)	84	86
F1-score (Neutral)	83	83
F1-score (Positive)	74	75

The comparison in Tabel 10 indicates that Logistic Regression slightly outperforms SVM in overall accuracy (82% vs. 81%). Additionally, Logistic Regression achieves higher recall for negative sentiment (90% vs. 86%), which is crucial in identifying customer dissatisfaction. It also performs slightly better in precision and F1-score for positive sentiment, suggesting a stronger ability to correctly classify positive feedback. However, SVM exhibits higher recall for neutral sentiment (92% vs. 89%), indicating that it is slightly more effective at detecting neutral reviews. Since neutral sentiment typically reflects uncertainty or mixed feedback, this aspect may be relevant depending on the business objectives.

Based on the results, Logistic Regression is recommended as the final sentiment classification model for Shopee Xpress. While both models demonstrate strong performance, Logistic Regression offers slightly better accuracy and recall for negative sentiment, making it more suitable for a business-oriented sentiment analysis system.

4 CONCLUSION

This study aimed to analyze customer sentiment toward Shopee Xpress services by employing Support Vector Machine (SVM) and Logistic Regression for sentiment classification. The analysis was conducted on 2480 customer reviews obtained from X and Google Play, categorizing them into positive, neutral, and negative sentiments. The findings reveal that negative sentiment dominates the dataset, with a significant proportion of reviews highlighting concerns regarding delivery delays and service inefficiencies. In contrast, positive sentiment primarily emphasizes fast delivery and ease of use, while neutral sentiment reflects factual or mixed opinions. Furthermore, sentiment distribution over time indicates a surge in negative sentiment during peak shopping periods, suggesting a correlation between increased demand and service challenges.

A comparative evaluation of the classification models indicates that Logistic Regression slightly outperforms SVM in overall accuracy (82% vs. 81%), particularly in detecting negative sentiment with higher recall. Meanwhile, SVM demonstrates a marginally better performance in identifying neutral sentiment. The confusion matrix analysis confirms that both models exhibit relatively low misclassification rates, reinforcing their reliability for sentiment classification tasks. These insights provide valuable implications for Shopee Xpress, emphasizing the need for targeted service improvements, particularly in addressing delivery reliability and customer support responsiveness. By leveraging sentiment analysis, Shopee Xpress can enhance operational strategies, mitigate service bottlenecks, and foster better customer engagement.

For future research, exploring deep learning-based approaches, such as transformer-based models, could further enhance classification accuracy and provide more nuanced sentiment interpretations. Additionally, integrating topic modeling techniques may offer deeper insights into specific aspects of customer concerns, facilitating more data-driven decision-making processes. Expanding the dataset to include a broader range of customer feedback sources or conducting cross-platform sentiment comparisons could also yield more comprehensive findings. Overall, this study underscores the importance of sentiment analysis in evaluating customer experiences and optimizing service quality in the e-commerce logistics sector.

AUTHOR'S CONTRIBUTION

Sewin Fathurrohman conceptualized the research framework, designed the methodology, and led the overall investigation. He was also responsible for dataset acquisition and model implementation. Irfan Ricky Afandi was responsible for data preprocessing, including cleaning and organizing the dataset for sentiment analysis. He also contributed to the implementation of machine learning models and conducted performance evaluations. Irma Wahyuningtyas conducted a comprehensive literature review, gathered relevant references, and ensured proper citation of previous works. She also played a key role in interpreting sentiment analysis results. Azis Styo Nugroho contributed to model evaluation and statistical analysis, deriving meaningful insights from the classification results. He also assisted in refining the discussion and aligning it with the research objectives. Firman Noor Hasan reviewed and edited the manuscript, ensuring coherence, clarity, and adherence to academic writing standards. He also verified the completeness and accuracy of references and citations. All authors contributed to the manual sentiment labeling process, discussed the findings, and reviewed the final manuscript before submission.

COMPETING INTERESTS

The authors declare that there are no competing interests (CI) or conflicts of interest (COI) related to this research. The study was conducted independently, without any external influence from organizations or entities that could affect the objectivity of the findings and conclusions.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all individuals who contributed to this research. Special appreciation is extended to those who provided valuable insights, constructive feedback, and support throughout the study. The authors also acknowledge the efforts involved in data collection, preprocessing, and analysis, which were crucial in ensuring the accuracy and reliability of the findings.

REFERENCES

[1] Y. Septiawan, A. Aglasia, and D. A. Muktiawan, "Peningkatan Performa Analisis Sentimen Ulasan Pelanggan terhadap Layanan Pengiriman Menggunakan Model Naïve Bayes yang Dioptimalkan dengan PSO," *ROUTERS J. Sist. dan Teknol. Inf.*, vol. 3, no. 1, pp. 10–19, 2025, doi: 10.25181/rt.v3i1.4001.

[2] N. A. Qadri, S. Alam, and M. Taufik, "Analisis Sentimen Brand Barenbliss Menggunakan Machine Learning (Studi Kasus : E-Commerce Shopee)," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 1–12, 2025, doi: https://doi.org/10.31004/innovative.v5i1.16163.

[3] T. Widarmanti and A. Munisera, "Analisis Sentimen Mengenai Kualitas Layanan Jasa Ekspedisi Shopee Xpress Pada Media Sosial Twitter," *J. Ilm. Univ. Batanghari Jambi*, vol. 23, no. 3, p. 3370, 2023, doi: 10.33087/jiubj.v23i3.4077.

[4] M. Taufiq, Recky, and Y. P. Sari, "Pengaruh Kompensasi, Motivasi, Dan Kepuasan Kerja Terhadap Kinerja Karyawan Pada Shopee Express Kabupaten Bandung," *J. Econ.*, vol. 12, no. 2, pp. 226–235, 2024, doi: https://doi.org/10.46750/economica.v12i2.297.

[5] I. R. Afandi, F. Noor Hasan, A. A. Rizki, N. Pratiwi, and Z. Halim, "Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes," *J. Linguist. Komputasional*, vol. 5, no. 2, pp. 63–70, 2022, doi: https://doi.org/10.26418/jlk.v5i2.107.

[6] M. Arsadhana, B. Efendi, and M. Trihudiyatmanto, "ANALISIS KEPUASAN PELANGGAN MELALUI SENTIMEN ULASAN MENGGUNAKAN ALGORITMA NAIVE BAYES," *J. Magisma*, vol. 13, no. 1, pp. 1–8, 2025, doi: https://doi.org/10.35829/magisma.v13i1.471.

[7] S. Nur Rismanah, R. Astuti, and F. M. Basysyar, "Penerapan Algoritma Support Vector Machine Dalam Menganalisis Sentimen Ulasan Pelanggan Shopeefood Berdasarkan Twitter," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 406–412, 2024, doi: 10.36040/jati.v8i1.8401.

[8] A. Simanungkalit, J. P. P. Naibaho, and A. De Kweldju, "Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Shopee Menggunakan Algoritma Naïve Bayes," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 659, 2024, doi: 10.35889/jutisi.v13i1.1826.

[9] A. M. Putri, W. K. Nofa, and D. A. P. Hapsari, "Penerapan Metode Bert Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Segari Di Google Play Store," *JUIT J. Ilm. Tek.*, vol. 4, no. 1, pp. 89–104, 2025, doi: https://doi.org/10.56127/juit.v4i1.1902.

[10] S. Wulandari, M. Trihudiyatmanto, and B. Efendi, "Analysis Of MSME Customer Perception Through Sentiment Reviews," *J. Magisma*, vol. 13, no. 1, pp. 48–55, 2025, doi: https://doi.org/10.35829/magisma.v13i1.476.

[11] Y. Handika, I. F. Hanif, and F. N. Hasan, "Analysis of Public Sentiment Towards POLRI's Performance using Naive Bayes and K-Nearest Neighbors," *IJID (International J. Informatics Dev.*, vol. 13, no. 1, pp. 386–399, 2024, doi: 10.14421/ijid.2024.4500.

[12] S. Heristian, M. Napiah, and W. Erawati, "Analisis Sentimen Ulasan Pelanggan Menggunakan Algoritma Naive Bayes pada Aplikasi Gojek," *Comput. Sci.*, vol. 5, no. 1, pp. 35–41, 2025, doi: https://doi.org/10.31294/coscience.v5i1.7775.

[13] S. N. Salombe, S. Alam, and V. Aris, "Analisis Sentimen Ulasan Netizen Pada Aplikasi Disney + Hotstar Di Aplikasi Google Play Store Menggunakan Machine Learning," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1–9, pp. 5968–5977, 2025, doi: https://doi.org/10.31004/innovative.v5i1.15846.

[14] Tukino, "Penerapan Algoritma Convolutional Neural Network Untuk Klasifikasi Sentimen Pada Layanan e-Commerce," *J. DESAIN DAN Anal. Teknol.*, vol. 4, no. 1, pp. 44–53, 2025, doi: https://doi.org/10.58520/jddat.v4i1.72.

[15] I. D. Yulisa, Gusmelia Testiana, and Imamulhakim Syahid Putra, "Classifying High School Scholarship Recipients Using the K-Nearest Neighbor Algorithm," *IJID (International J. Informatics Dev.*, vol. 11, no. 2, pp. 232–241, 2023, doi: 10.14421/ijid.2022.3804.

[16] A. G. Budianto, Rusilawati, A. T. E. Suryo, G. R. Cahyono, A. F. Zulkarnain, and Martunus, "Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android," *J. Sains dan Inform.*, vol. 10, no. 2, pp. 1–10, 2024, doi: 10.34128/jsi.v10i2.911.

[17] N. Sari, M. Jazman, T. K. Ahsyar, Syaifullah, and A. Marsal, "Penerapan Algoritma Klasifikasi Naive Bayes dan Support Vector Machine untuk Analisis Sentimen Cyberbullying Bilingual di Aplikasi X," *Sist. J. Sist. Inf.*, vol. 14, no. 1, pp. 211–224, 2025, doi: https://doi.org/10.32520/stmsi.v14i1.4799.

[18] T. G. W. M. Sidabutar and D. Juardi, "Analisis Sentimen Masyarakat Terhadap Penggunaan Halodoc Sebagai Layanan Telemedicine Di Indonesia," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 571–575, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i1.5682.

[19] B. Restya and N. Cahyono, "Optimasi Metode Klasifikasi Menggunakan FastText dan Grid Search pada Aanalisi Sentimen Ulasan Aplikasi SeaBank," *JIKO (Jurnal Inform. dan Komputer)*, vol. 9, no. 1, pp. 226–238, 2025, doi: 10.26798/jiko.v9i1.1523.

[20] Z. B. Toyibah, Y. N. Putri, Puandini, Z. M. Widodo, and A. Tsalitsatun, "Perbandingan Kinerja Algoritma Multinomial Naïve Bayes dan Logistic Regression pada Analisis Sentimen Movie Ratings IMDB," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 10, no. 2, pp. 181–189, 2024, doi: https://doi.org/10.21107/edutic.v10i2.28150.

[21] A. R. Hidayati, A. S. Fitrani, and M. A. Rosid, "Analisa Sentimen Pemilu 2019 Pada Judul Berita Online Menggunakan Metode Logistic Regression," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 4, no. 2, pp. 298–305, 2023, doi: https://doi.org/10.30645/kesatria.v4i2.164.

[22] E. G. Radjah and A. C. Talakua, "Analisis Sentimen Komentar Terhadap Konten Tenun NTT di Youtube Menggunakan Metode SMOTE dan Logistic Regression," *J. Transform.*, vol. 13, no. 2, pp. 84–94, 2024, doi: https://doi.org/10.58300/transformatif.v13i2.1005.

[23] N. A. Amalia, I. T. Utami, and Y. Wilandari, "Analisis Sentimen Kebijakan Penyelenggara Sistem Elektronik Lingkup Privat Menggunakan Penalized Logistic Regression Dan Support Vector Machine," *J. Gaussian*, vol. 12, no. 4, pp. 560–569, 2024, doi: 10.14710/j.gauss.12.4.560-569.

[24] M. I. Burhan, "Analisis Sentimen Publik Terhadap Aplikasi Mypertamina Pada Google Playstore," *J. Manaj. Perbank. Keuang. Nitro*, vol. 8, no. 1, pp. 1–12, 2025, doi: 10.56858/jmpkn.v8i1.390.

[25] B. Kholifah, I. Thoib, N. Sururi, and N. D. Kurnia, "Analisis Sentimen Warganet Terhadap Isu Layanan Transportasi Online Berbasis InSet Lexicon Menggunakan Logistic Regression," *Kumpul. J. Ilmu Komput.*, vol. 11, no. 1, pp. 14–25, 2024, doi: http://dx.doi.org/10.20527/klik.v11i1.655.

[26] D. Prasetia, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, "Analisis Sentimen Pengguna Aplikasi Mybluebird Dengan Algoritma Naïve Bayes Di Playstore," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 602–607, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i1.5687.

[27] B. A. Putra *et al.*, "Optimisasi Algoritma K-Means Dengan Metode Reduksi Dimensi Untuk Pengelompokan Big Data Dalam Arsitektur Cloud Computing," *J. Softw. Eng. Inf. Syst.*, vol. 5, no. 1, pp. 1–8, 2025, doi: https://doi.org/10.37859/seis.v5i1.7616.

[28] R. D. R. Apriliansyah, R. AstutI, W. Prihartono, and R. Hamonangan, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pengunjung Di Pantai Kejawanan," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 917–926, 2025,

doi: http://dx.doi.org/10.23960/jitet.v13i1.5774.

[29] S. Setyabudi and E. Aryanny, "Sentiment Analysis Of Lazada Marketplace User Ratings With Naïve Bayes And Support Vector Machine Methods," *J. INOVTEK POLBENG - SERI Inform.*, vol. 10, no. 1, pp. 422–433, 2025, doi: https://doi.org/10.35314/sww8cg21.

[30] R. R. Rismansyah, A. Sudiarjo, and T. Mufizar, "Analisis Sentimen Ulasan Shopee Pada Google Play Store Menggunakan Algoritma Naive Bayes," *J. Elektro Inform. Swadharma*, vol. 5, no. 1, pp. 109–119, 2025, doi: https://doi.org/10.56486/jeis.vol5no1.661.

[31] S. Fathurrohman, I. R. Afandi, and F. N. Hasan, "Sentiment Analysis of TIMNAS Indonesia ' s Participation in the Asian Cup U23 2024 on X Using Naive Bayes and SVM," *IJID (International J. Informatics Dev.*, vol. 13, no. 1, pp. 434–447, 2024, doi: 10.14421/ijid.2024.4504.

[32] Y. N. Fernando, A. Y. Rahman, and R. P. Putra, "Analisis Sentimen pada Ulasan Konsumen Ayam Goreng Nelongso di Google Maps Menggunakan Metode Bidirectional Long Short-Term Memory (Bi-LSTM)," *J. Ilm. Mhs. Sains UNISMA Malang*, vol. 3, no. 1, pp. 65–85, 2025, doi: http://dx.doi.org/10.33474/jimsum.v3i1.26592.

[33] F. Andriansyah and P. Astuti, "Analisis Sentimen Komentar Konten Edukatif Di Instagram Dengan Metode Naïve Bayes Dan Support Vector Machine," *J. Imtechno*, vol. 6, no. 2, pp. 27–31, 2025, doi: https://doi.org/10.31294/imtechno.v6i1.7262.

[34] M. F. A. Shidiq and D. Alita, "Analisis Sentimen Masyarakat Terhadap Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan SVM," *J. Sist. Inf. dan Inform.*, vol. 8, no. 1, pp. 24–35, 2025, doi: https://doi.org/10.47080/simika.v8i1.3624.

[35] H. T. Saputra, A. Damanik, M. M. Shaquille, M. A. Rusydi, and M. Habib, "Analisis Modelling Pada Reviewes Lazada Indonesia Menggunakan Latent Dirichlet Allocation (LDA) Untuk Optimalisasi Strategi Bisnis," *JEKIN J. Tek. Inform.*, vol. 5, no. 1, pp. 361–370, 2025, doi: https://doi.org/10.58794/jekin.v5i1.1325.

[36] W. Hadi, T. D. Putra, and D. Oktafiani, "Analisis Sentimen Terhadap Review Produk Lazada Indonesia Menggunakan Random Forest Classifier," *Innov. J. Soc. Sci. Res.*, vol. 5, no. 1, pp. 2348–2357, 2025, doi: https://doi.org/10.31004/innovative.v5i1.17934.

[37] K. Khotimah, Martanto, A. R. Dikananda, and A. Rifa'i, "Analisis Sentimen Ulasan Aplikasi Pintu Di Google Play Store Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 963–969, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i1.5789.

[38] E. D. K. Wardani, F. F. Yo, and W. N. Meylugita, "Implementasi Algoritma Naïve Bayes Untuk Analisis Ulasan Pengguna Untuk Aplikasi Seabank Di Google Play Store," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 1, pp. 13–24, 2025, doi: https://doi.org/10.69916/jkbti.v4i1.193.

[39] D. Fadila and M. Ikhsan, "Analisis Sentimen Pada Aplikasi Tokopedia Menggunakan Metode Support Vector Machine," *Progresif J. Ilm. Komput.*, vol. 21, no. 1, pp. 238–245, 2025, doi: 10.35889/progresif.v21i1.2593.

[40] R. Maheri, F. N. Salisah, and F. Muttakin, "Analisis Sentimen Ulasan Aplikasi M-Paspor Menggunakan NAIVE BAYES DAN SUPPORT VECTOR MACHINE," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 448–458, 2025, doi: https://doi.org/10.29100/jipi.v10i1.5826.

[41] R. Yusuf, K. Bahumatra, N. Komaria, E. A. Aqma, and L. Cahyani, "Analisis Sentimen Terhadap Aplikasi Google Meet Berdasarkan Komentar Pengguna Menggunakan Metode Logistic Regresion," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 11, no. 1, pp. 53–64, 2024, doi: https://doi.org/10.21107/edutic.v11i1.28113.

[42] I. R. Afandi, I. F. Hanif, F. N. Hasan, E. S. Sinduningrum, Z. Halim, and N. Nunik, "Analisis Sentimen Opini Masyarakat Terkait Penyelenggaraan Sistem Elektronik Menggunakan Metode Logistic Regression," *J. Linguist. Komputasional*, vol. 5, no. 2, p. 77, 2022, [Online]. Available: https://t.co/23c4krbjp

[43] A. N. Sativa, A. Rizky, I. Putri, J. A. Putri, A. Irsyad, and Islamiyah, "Analisis Sentimen Twitter Ibu Kota Negara Nusantara Menggunakan Algoritma Naive Bayes , Logistic Regression dan K-Nearest Neighbors," *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 34–40, 2024, doi: https://doi.org/10.30872/atasi.v3i2.1371.

[44] T. Prasetyo, A. A. Waskita, and T. Taryo, "Analisis Sentimen Pengguna Seputar Kendaraan Listrik Di Twitter Dengan Penerapan Algoritma Naïve Bayes , KNN , dan Decision Tree untuk Klasifikasi," *J. Sist. Komput. dan Kecerdasan Buatan*, vol. 8, no. 2, pp. 108–116, 2025, doi: https://doi.org/10.47970/siskom-kb.v8i1.783.

[45] I. R. Afandi, S. Fathurrohman, and M. W. A. Wibowo, "Sentiment Analysis of Mandatory Halal Certification Policy on Twitter Using the Naive Bayes and K-Nearest Neighbors Algorithm," *Halal Res.*, vol. 5, no. 1, pp. 37–58, 2025, doi: https://doi.org/10.12962/j22759970.v5i1.1806.

[46] F. M. Hidayat and H. Sanjaya, "Analisis Sentimen Publik Terhadap Penjualan Iphone 16 Dan Kebijakan TKDN Di Indonesia," *Infotech J.*, vol. 11, no. 1, pp. 73–80, 2025, doi: https://doi.org/10.31949/infotech.v11i1.13159.

[47] P. W. Hardjita, Nurochman, and R. Hidayat, "Sentiment Analysis of Tweets on Prakerja Card using Convolutional Neural Network and Naive Bayes," *IJID (International J. Informatics Dev.*, vol. 10, no. 2, pp. 82–91, 2022, doi: 10.14421/ijid.2021.3007.

[48] A. Septini, Susanto, and Elmayati, "Analisis Sentimen Masyarakat di Twitter Mengenai Open AI CHATGPT Menggunakan Metode Support Vector Machine (SVM)," *Bull. Comput. Sci. Res.*, vol. 5, no. 2, pp. 138–149, 2025, doi: 10.47065/bulletincsr.v5i2.475.

[49] M. Z. Ramadhan and R. Mubarak, "Analisis Sentimen Masyarakat Terhadap Kenaikan Bahan Bakar Minyak Pada Media Sosial Youtube Dengan Metode K-Nearest Neighbor Dan Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 984–991, 2025, doi: https://doi.org/10.36040/jati.v9i1.12526.

[50] A. Saninah, W. Prihartono, C. L. Rohmat, and K. Cirebon, "Analisis Sentimen Pengguna Terhadap Aplikasi Duolingo Dengan Naïve Bayes Classifier," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 619–627, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i1.5691.

[51] Alvinalia, A. Rahim, and T. A. Y. Siswa, "Analisis Sentimen Aplikasi Mysiloam Menggunakan Metode Naive Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 1, pp. 1546–1555, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i1.5997.

[52] Y. R. Z. Mustopo and Afiyati, "Analisis Sentimen Proyek Strategis Nasional Food Estate Menggunakan Algoritma Naïve Bayes, Logistic Regression dan Support Vector Machine," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 485–494, 2025, doi: https://doi.org/10.35870/jtik.v9i2.3312.