



**Equivalent: Jurnal Ilmiah Sosial Teknik**  
**Volume 8, Issue 2, 531-548**  
e\_ISSN: 2775-0833  
<https://jurnalequivalent.id/index.php/jequi/index>  
DOI: [doi.org/10.59261/jequi.v8i2.324](https://doi.org/10.59261/jequi.v8i2.324)

## **Sentiment Analysis Comparison of Two E-Commerce Platforms Using Random Forest, Support Vector Machine, Logistic Regression, and IndoBERT**

**Theresia Vania Davita  
Suyana<sup>1\*</sup>**

Universitas Bina Nusantara,  
Indonesia

**Sfenrianto<sup>2</sup>**

Universitas Bina Nusantara,  
Indonesia

---

**\*Corresponding author:**

Theresia Vania Davita Suyana, Universitas  
Bina Nusantara, Indonesia.

✉[Theresia.suyana@binus.ac.id](mailto:Theresia.suyana@binus.ac.id)

---

**Article Info:**

**Article history:**

Received: April 21, 2026

Revised: June 06, 2026

Accepted: June 08, 2026

---

**Keywords:**

Sentiment Analysis; IndoBERT;  
Machine Learning; Deep Learning;  
E-Commerce; VADER-Based

---

**Abstract**

**Background:** The rapid growth of e-commerce in Indonesia has generated massive volumes of user-generated reviews. A critical mismatch exists between numerical star ratings and the actual sentiment expressed in review texts, creating unreliable signals for platform management and highlighting the need for text-based sentiment analysis.

**Objective:** This study aims to analyze user review sentiments towards two leading e-commerce platforms in Indonesia using Machine Learning and Deep Learning approaches.

**Methods:** The analysis process was conducted through the CRISP-DM stages, including data cleaning, labeling, model training, and performance evaluation. Two types of labeling were used, namely Rating-Based and VADER-Based, to compare the accuracy levels of sentiment classification. Four models were applied: Random Forest, Support Vector Machine (SVM), Logistic Regression, and *IndoBERT*. VADER labeling was adapted for Indonesian through preprocessing with an Indonesian-English translation layer.

**Results:** Based on the evaluation results, the *IndoBERT* model showed the best performance on e-commerce X, with an accuracy of 0.96 using VADER-based labeling. Meanwhile, for e-commerce Y, Random Forest achieved an accuracy of 0.81 using VADER labeling. These results indicate that *IndoBERT*'s Transformer architecture with contextual embeddings enabled superior understanding of Indonesian semantic nuances. Random Forest's advantage on e-commerce Y (627 samples per label) reflects a lower overfitting risk compared to deep learning models on small datasets.

**Conclusion:** This study demonstrates the effectiveness of combining *IndoBERT* and VADER in Indonesian sentiment analysis and can serve as a reference for the e-commerce industry to improve service quality and customer satisfaction strategies.

---

**To cite this article:** Suyana, T. F. D., & Sfenrianto. (2026). Sentiment Analysis Comparison of Two E-Commerce Platforms Using Random Forest, Support Vector Machine, Logistic Regression, and IndoBERT. *Equivalent: Jurnal Ilmiah Sosial Teknik*, 8(2), 531-548. <https://doi.org/10.59261/jequi.v8i2.324>

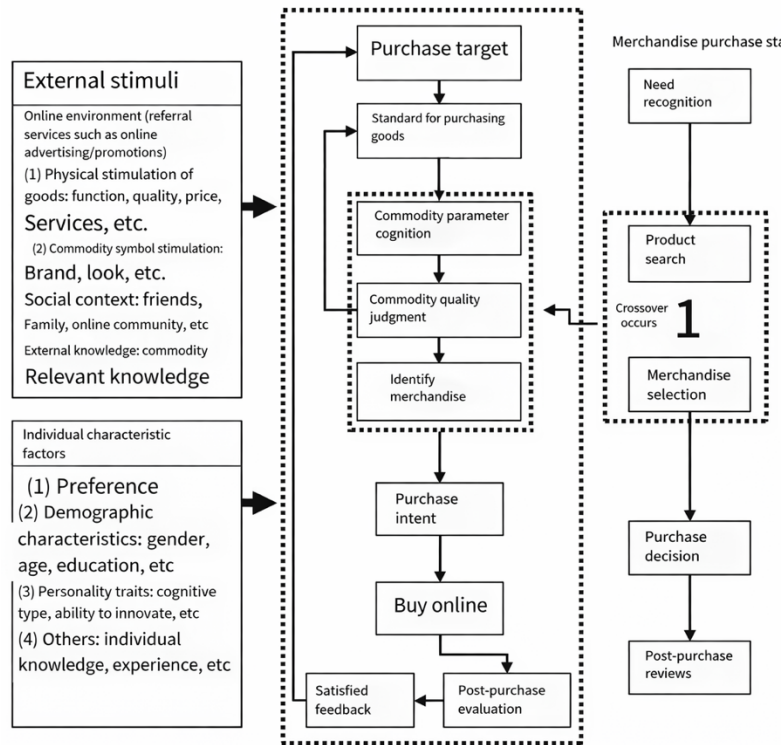
---

### **INTRODUCTION**

In this era, online shopping platforms, or e-commerce platforms, are available in various forms. Some of them include X and Y. Sometimes, online shopping is also referred to as an internet store. The term online shopping platform is defined as the convenience of shopping online, which has increased its popularity by allowing consumers to explore and purchase products from the comfort of their homes. This transformation has reshaped the retail sector, providing a wide range

of goods and services accessible through digital channels (Gallin & Portes, 2024). Since the introduction of the Internet, online shopping has gained its own prominence. The growth of digital-era platforms has contributed to developments in customer behavior. Customers adapt to changes that occur, along with evolving needs, and also follow trends, which have become a competitive area for both platforms, X and Y. Most applications utilize big data analytics to provide customers with recommended products based on their preferences or search history.

Basic information about how online shopping works is essentially that buyers select their desired items within the application and then place an order; the order will subsequently be delivered. Customer decisions are not always linear; there is often a stage where they revisit previous choices before making a final purchase (Liu, 2024). At present, a model of online customer product selection decisions has been developed.



**Figure 1.** Model of Online Consumer Product's Selection Decision-Making  
Source: research data

Figure 1 shows the model of how customers choose the products they want to buy. Consumer product selection is a critical aspect of purchasing behavior, influenced by various factors such as personal preferences, brand perception, price, quality, and external recommendations. Understanding how consumers choose products can help businesses improve their marketing strategies, increase customer satisfaction, and enhance product offerings.

Over the years, different models have been developed to explain consumer decisions in product selection. These models incorporate psychological, social, and economic factors that drive purchasing decisions. Some of the most commonly used models include the rational choice model, the behavioral model, and the Multi-Attribute Decision Model (Alzate et al., 2022; Liang et al., 2021).

Another study entitled "A Study on Consumer Behaviour Towards Online Shopping" explores how the internet has transformed traditional retail by providing opportunities for consumers not only to buy products online but also to compare prices, features, and community support with traditional retail. Businesses can leverage online platforms to reduce marketing costs, gather consumer feedback, and conduct satisfaction surveys, thereby increasing their competitiveness in the digital market (Santhosh & Basavarajappa, 2022). However, one consideration is shipping costs; online retail often becomes more expensive due to shipping fees.

Another research study entitled "Online and Offline Shopping: Decision-Making Factors that Influence Consumer Purchases" highlights how e-commerce has significantly reshaped the retail industry by offering consumers the ability to browse and purchase products online easily while providing tools to compare prices, assess product features, and evaluate after-sales services (Schulze, 2020; Santhosh & Basavarajappa, 2022). This study emphasizes that the rise of online shopping benefits customers seeking convenience and better deals while also providing advantages to businesses looking to reduce operational costs, improve marketing efficiency, and enhance customer engagement through digital platforms. By leveraging the online marketplace, companies can interact directly with customers, gather valuable feedback, and conduct satisfaction surveys, which helps refine their services and improve the customer experience. This shift toward digital retail has enabled businesses to remain competitive in a market increasingly driven by technology, where user experience, competitive pricing, and effective customer service play crucial roles in consumer decision-making.

In conclusion, factors influencing consumer purchasing decisions such as product prices, discount availability, delivery efficiency, and customer service quality can significantly affect the competitive position of e-commerce platforms like Y and X. If one platform offers better deals, faster shipping, or superior customer support, it is more likely to attract and retain a larger customer base, thereby gaining an advantage over competitors. Furthermore, businesses that effectively leverage customer insights by analyzing purchasing behavior and feedback can develop strategies that enhance their market presence and foster customer loyalty. Therefore, the ability to understand and adapt to consumer preferences is crucial for X and Y to maintain their dominance in the competitive e-commerce landscape.

The rapid evolution of e-commerce has significantly transformed the retail landscape, with major players like Amazon setting new standards in consumer expectations. This phenomenon, often called the "Amazon Effect," encompasses various impacts, including increased pricing flexibility, uniform pricing, and the need for traditional retailers to enhance in-store experiences to remain competitive (Ibrahim, 2023). Research has shown that the transparency and convenience offered by online platforms have driven physical stores to adopt more dynamic pricing strategies and integrate advanced technologies to streamline the shopping experience. For example, retailers now focus on creating engaging store environments and leveraging data analytics to better understand consumer preferences. Moreover, competitive pressure from major e-commerce companies has led to a re-evaluation of supply chain efficiency and customer service protocols among traditional retailers. This shift not only changes pricing dynamics but also influences consumer behavior, with shoppers increasingly expecting seamless and personalized experiences across all retail channels.

In the context of Indonesian e-commerce platforms like Y and X, these global trends have significant implications. To maintain and enhance their competitive positions, these platforms must continue to innovate to meet evolving consumer expectations. This includes implementing flexible pricing models, ensuring user-friendly and engaging online shopping experiences, and optimizing logistics for faster delivery times. Furthermore, as consumers become more accustomed to the convenience and transparency offered by leading global e-commerce players, Y and X need to invest in advanced technologies, such as artificial intelligence and data analytics, to personalize shopping experiences and anticipate market trends. By aligning their strategies with these evolving trends, Y and X can strengthen their market presence and foster greater customer loyalty in the competitive e-commerce landscape. In the rapidly growing digital market, online shopping has become an integral part of consumer behavior, driven by convenience, competitive prices, and diverse product choices. E-commerce platforms like Y and X have experienced significant growth, serving millions of users across Southeast Asia. However, with increasing competition and changing consumer expectations, understanding the factors influencing consumer preferences and purchasing decisions is crucial.

Despite their success, Y and X face challenges in maintaining customer loyalty, optimizing user experience, and differentiating themselves in a highly competitive market. Factors such as pricing strategies, platform usability, delivery efficiency, and customer service quality play essential roles in shaping consumer perceptions. Additionally, emerging market trends, such as personalized recommendations, AI-powered shopping experiences, and digital payment

innovations, are increasingly impacting consumer choices and the competitive positioning of these platforms.

Given these challenges, it is important to analyze how consumer behavior, sentiment, and market trends influence the competitive dynamics between Y and X. Identifying key factors driving customer satisfaction and loyalty can provide valuable insights for businesses to refine their strategies and enhance their market presence.

In Indonesia's rapidly growing e-commerce sector, platforms like X and Y have emerged as dominant players, each implementing various strategies to attract consumers. A study entitled "Comparative Analysis of Promotion Quality, Product Diversity, and Ease of Use Between E-Commerce X and Y" investigated whether there are significant differences between these two platforms in terms of promotion quality, product diversity, and ease of use. This study used a comparative quantitative approach, utilizing questionnaires distributed to 300 respondents 150 for each platform. The findings revealed that only the promotion quality variable showed a significant difference between X and Y, with a significance value of 0.04, which is less than the threshold of 0.05. In contrast, the product diversity and ease of use variables did not show significant differences, with significance values of 0.326 and 0.432, respectively, both exceeding the 0.05 threshold (Glori & Widjaja, 2024).

These results indicate that while both platforms offer comparable product variety and user-friendliness, their promotional strategies differ significantly. This difference in promotion quality can influence consumer preferences and purchasing decisions, potentially affecting the competitive dynamics between X and Y. Understanding these differences is crucial for both platforms as they seek to enhance their market positions and better meet consumer needs in Indonesia's evolving e-commerce landscape. Ensuring customer trust is also essential to maintain continued purchases on their applications.

Based on 35 important Indonesian e-commerce statistics covering both X and Y data, in the fourth quarter of 2024, X recorded very positive performance with a total of 3 billion gross orders, a 20% increase compared to the same period the previous year, and a Gross Merchandise Value (GMV) reaching \$28.6 billion, an increase of 23.5%. Throughout 2024, X successfully obtained a total of 10.9 billion orders with a GMV of \$100.5 billion, a significant increase compared to 2023, which only recorded 8.2 billion orders and a GMV of \$78.5 billion. Local product sales also surged by up to 200% thanks to the export strategy and promotions on the X Pilih Lokal channel. In terms of visits, as of February 2025, X received 152.6 million visits with 53.2 million unique visitors, an average visit duration of 9 minutes 26 seconds, and 69.1% came from direct visits. The majority of visitors are male (56.92%), with the dominant age group being 25–34 years (34.06%), and the majority of access comes from mobile web at 58.58%.

Conversely, Y experienced a decline in sales in the first half of 2024, achieving only Rp5.2 trillion, lower than Rp6.8 trillion in the same period the previous year. Additionally, visits to the Y site also decreased during the fourth quarter of 2024 by 8.6%, from 73.4 million visits in October to 67.1 million in December. Nevertheless, the number of sellers increased from 12 million in 2022 to 14 million in 2023. As of February 2025, Y recorded 61.6 million visits with 21 million unique visitors, an average visit duration of 11 minutes 28 seconds, and 57.38% of total visits came from direct access. As many as 60.43% of visitors are male, with the 25–34-year age group dominating at 34.60%, and access devices dominated by desktop at 54.84%. This data shows that although X is superior in terms of sales performance and traffic, Y still shows potential through the number of sellers and longer visit duration.

The selection of Platforms X and Y is based on their strategic positions in Indonesia's digital economy. Based on statistical data from the fourth quarter of 2024, Platform X shows significant market dominance with a GMV reaching \$100.5 billion and 152.6 million monthly visits. On the other hand, Platform Y, despite experiencing sales fluctuations, still shows great potential with longer user visit durations and a seller base reaching 14 million. The contrast in business performance between X, which is growing rapidly, and Y, which is undergoing market adjustments, makes them ideal research objects to analyze how the dynamics of user sentiment are reflected in text reviews on each platform.

Besides financial performance, statistical visit data also shows a high level of user dependence and trust in these two brands. It was recorded that 69.1% of visits to Platform X and 57.38% of visits to Platform Y came from direct traffic. From an Information Systems Management perspective, this high direct traffic indicates that the majority of users no longer search via search engines but directly access the platform because they already have loyalty or dependence on the service.

**Table 1.** Evidence of Inappropriate Reviews on E-Commerce X

<b>Review on E-Commerce X</b>	<b>Rating</b>
unclear often don't get driver expensive shipping	5
Shipping is very very slow, sluggish...	5
Check out 8 times this.. delivery almost a week just arrived...	
Automatically switch allegiance like this	
good job	4
once	3
Very helpful Ma'am	2

Source: research data

The phenomenon of mismatch between review text and star ratings in Table 1 demonstrates that quantitative ratings often fail to reflect actual user satisfaction. For example, reviews containing critical complaints regarding operations, such as "unclear often don't get driver expensive shipping," and disappointment with logistics, like "shipping is very very slow, sluggish...", are still accompanied by a rating of 5. Conversely, there is also a low-rating bias, where an explicitly positive review, such as "Very helpful Ma'am," is only given a rating of 2. This inconsistency highlights a significant information gap, in which star numbers no longer accurately represent users' genuine emotions. Therefore, applying text-based sentiment analysis through automatic labeling methods, such as VADER, becomes a much more precise and objective instrument for e-commerce management to capture the reality of customer experience.

**Table 2.** Evidence of Inappropriate Reviews on E-Commerce Y

<b>Review on E-Commerce Y</b>	<b>Rating</b>
The latest update has more and more errors, failing to make payments, transaction cancellations, and getting slower.	5
Cool fast and smooth	4
Please ask Tokopedia to re-examine sellers who sell products that do not match their products because there are many fraudulent sellers selling at low prices indicating fraud and sending goods that do not match the advertisement. many victims and please Tokopedia also help those victims don't complicate, for example I am a victim of fraud at the Ayu Wijaya28 Shop, please blacklist the Ayu Wijaya28 Shop immediately and investigate the clear address of the shop. Thank You	5
Bad app	5
Good service & quick	4

Source: research data

Similar to the findings on the previous platform, data in Table 2 shows that E-Commerce Y also experiences serious problems regarding the validity of quantitative ratings. The most prominent business issue is the rating bias on critical complaints, where users still give a Rating of 5 even though the review narrative reports systemic failures, such as errors in application updates, payment transaction failures, and indications of fraud by sellers. This indicates user behavior of intentionally giving high stars so that their reports of losses or complaints are more visible to management or other users. Additionally, the emergence of positive reviews that are instead assigned lower ratings (Rating 4) further confirms that star numbers are highly subjective. Managerially, this finding demonstrates that relying solely on ratings will provide misleading information for Platform Y in diagnosing the causes of its business performance decline. Therefore, using text-based sentiment analysis (VADER) becomes a crucial solution to capture the reality of customer experience both honestly and accurately.

Conducting sentiment analysis is thus essential as an instrument for Brand Reputation Management. Considering the large portion of loyal (direct) users, the emergence of undetected negative sentiment can pose fatal risks to user trust. If the user experience is disrupted, as might be reflected in the fluctuation of Platform Y's visit data, which fell by 8.6%, text-based sentiment analysis is the only way to deeply understand users' technical and non-technical complaints. By accurately identifying positive, negative, and neutral sentiments, Platform X can maintain its market dominance, while Platform Y can formulate system improvement strategies to halt traffic decline and enhance its sales performance.

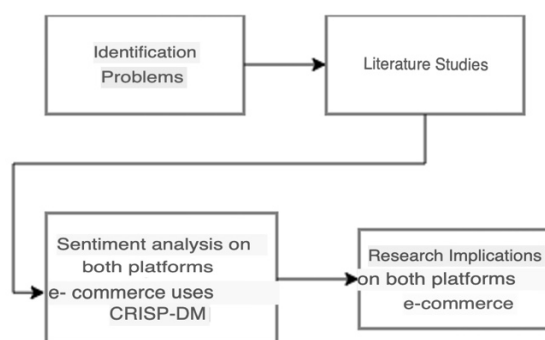
This research is based on sentiment analysis, which has strong theoretical grounding in the field of information systems. Previously, it was discussed that decision support systems can influence consumer behavior in both shopping and review activities. Sentiment analysis itself employs machine learning using the Python programming language. Machine learning can help process data into actionable information to support decision-making in e-commerce companies (Ait et al., 2024). Decision-making in e-commerce platforms X and Y considers aspects of consumer behavior, trends, and marketing strategies.

The high number of reviews mentioned above is not unique to a single e-commerce platform. Data from X and Y indicate that fluctuations can occur across all e-commerce applications. Existing research focuses primarily on sentiment analysis in a single e-commerce platform (Novitasari et al., 2024), while this study uses two e-commerce platforms to compare evaluation results using the Random Forest, Support Vector Machine (SVM), and Logistic Regression approaches. Based on literature studies by Rayhan, Wijoyo, and Putra (2024) and Ananta Firdaus et al. (2024), the Random Forest algorithm exhibits stable results with accuracy and other metrics above 80%, which is considered robust for text classification. Literature studies by Idris et al. (2023) and Sanjaya et al. (2023) indicate that the SVM algorithm achieves high accuracy, around 90%, and is well-suited for text classification, where review content is primarily textual. Additionally, Novirianto (2023) shows that Logistic Regression also produces reliable results comparable to Random Forest.

The benefit of this research lies in conducting sentiment analysis on consumer feedback for X and Y, which provides several valuable insights. The primary objective of this sentiment analysis is to compare customer perceptions of X and Y, the two largest e-commerce platforms in Southeast Asia.

## METHOD

Data were collected through web scraping of user reviews from the Google Play Store for both Platform X (Shopee) and Platform Y (Tokopedia) during the period January–December 2024. A total of 12,252 reviews were collected for Platform X and 1,881 reviews for Platform Y prior to preprocessing. Inclusion criteria consisted of reviews written in the Indonesian language with a minimum of five characters. The scraping was conducted using the google-play-scraper Python library. The research stages can also be conceptualized as a conceptual framework. While the framework generally followed the same structure, its interpretation could vary depending on the specific research context.



**Figure 2.** Conceptual Framework

Source: research data



The word cloud results of Logistic Regression on E-Commerce Y based on VADER in Figure 3 show that the words "free," "shipping cost," and "tokopedia" are most dominant in positive reviews, indicating user satisfaction with promotions and shipping convenience. In negative reviews, the words "error," "application," and "goods" appear most prominently, reflecting complaints related to technical application issues or product problems. Meanwhile, the words "tokopedia," "application," and "good" dominate neutral reviews, suggesting that users often mention the platform and its functions without expressing strong emotions. Overall, the Logistic Regression model effectively maps thematic differences between positive sentiments, which focus on promotional benefits, and negative sentiments, which are dominated by technical application constraints.

**Table 3.** Top 20 Word Frequency for E-Commerce Y on Logistic Regression based on VADER

No	Positive Word	Frequency	Negative Word	Frequency	Neutral Word	Frequency
1	<i>gratis</i> (free)	25	<i>error</i> (error)	30	Tokopedia	16
2	<i>ongkir</i> (shipping fee)	23	<i>barang</i> (item)	26	aplikasi (application)	13
3	good	13	Tokopedia	25	<i>mantap</i> (great)	12
4	Tokopedia	13	<i>aplikasi</i> (application)	22	<i>bagus</i> (good)	11
5	<i>nya</i> (-its/-the")	12	<i>nya</i> (its/the)	22	<i>nya</i> (its/the)	11
6	<i>promo</i> (promotion)	9	<i>gak</i> (not/no)	20	Tokopedia	8
7	top	9	<i>tolong</i> (please)	19	<i>gak</i> (not/no)	7
8	care	9	tokped (Tokopedia slang)	17	<i>membantu</i> (helpful)	6
9	<i>belanja</i> (shopping)	7	<i>aja</i> (just/only)	16	<i>uang</i> (money)	6
10	tokped (Tokopedia slang)	7	dana (funds)	14	<i>sistem</i> (system)	5
11	<i>mahal</i> (expensive)	6	<i>udah</i> (already)	14	<i>belanja</i> (shopping)	5
12	<i>ongkirnya</i> (shipping cost)	6	<i>akun</i> (account)	13	<i>pembayaran</i> (payment)	5
13	<i>voucher</i> (voucher)	6	<i>belanja</i> (shopping)	12	<i>toko</i> (store)	5
14	<i>toko</i> (store)	6	<i>kalo</i> (if)	12	<i>sih</i> (anyway)	5
15	best (best)	6	<i>pengiriman</i> (delivery)	11	<i>mudah</i> (easy)	5
16	<i>gak</i> (not/no)	5	<i>cancel</i> (cancel)	10	<i>banget</i> (very)	5
17	toped (Tokopedia slang)	5	<i>penjual</i> (seller)	10	<i>penjual</i> (seller)	5
18	<i>aplikasi</i> (application)	4	<i>pas</i> (just right)	10	<i>transaksi</i> (transaction)	4



No	Positive Word	Frequency	Negative Word	Frequency	Neutral Word	Frequency
4	good	13	<i>gak</i> (not/no)	21	<i>bagus</i> (good)	8
5	<i>belanja</i> (shopping)	11	Tokopedia	19	<i>barang</i> (item)	7
6	<i>voucher</i> (voucher)	10	<i>aja</i> (just/only)	19	<i>aplikasi</i> (application)	6
7	<i>gak</i> (not/no)	10	<i>tolong</i> (please)	17	<i>membantu</i> (helpful)	6
8	<i>promo</i> (promotion)	9	<i>nya</i> (its/the)	17	<i>belanja</i> (shopping)	5
9	<i>aplikasi</i> (application)	9	tokped (Tokopedia slang)	16	<i>pembayaram</i> (payment)	5
10	<i>aja</i> (just/only)	9	<i>udah</i> (already)	14	<i>mudah</i> (easy)	5
11	<i>nya</i> (its/the)	9	<i>akun</i> (account)	13	<i>transaksi</i> (transaction)	4
12	top	8	<i>pengiriman</i> (delivery)	10	<i>banget</i> (very)	4
13	tokped (Tokopedia slang)	8	<i>bayar</i> (pay)	10	<i>penjual</i> (seller)	4
14	<i>ongkirnya</i> (shipping cost)	8	<i>belanja</i> (shopping)	10	<i>terbaik</i> (best)	3
15	<i>mahal</i> (expensive)	6	<i>kalo</i> (if)	10	<i>susah</i> (difficult)	3
16	<i>diskon</i> (discount)	6	dana (funds)	9	<i>kasih</i> (give)	3
17	Shopee	6	cancel	9	<i>pengiriman</i> (delivery)	3
18	<i>banget</i> (very)	6	<i>sistem</i> (system)	9	<i>cepat</i> (fast)	3
19	<i>toko</i> (store)	6	bgt (very slang)	9	<i>gak</i> (not/no)	3
20	best	6	<i>pas</i> (right)	9	<i>uang</i> (money)	3

Source: research data

Table 4 shows the Top 20 Word Frequency results of VADER-based IndoBERT analysis for e-commerce Y, which aligns with the word cloud visualization. In the positive category, the words "free," "shipping cost," and "tokopedia" dominate, indicating user appreciation for promotions and free shipping services. Meanwhile, the words "error," "application," and "goods" dominate the negative category, highlighting technical problems and unsatisfactory shopping experiences. For the neutral category, words such as "great," "tokopedia," and "good" appear most frequently, representing reviews that are descriptive without strong emotional polarity. Overall, these results confirm that IndoBERT effectively groups words according to sentiment polarity, with the context of words in each category consistent with user perceptions of e-commerce Y services.

After analyzing all the evaluation results for e-commerce X and Y, in terms of rating-based and VADER-based accuracy for e-commerce X, IndoBERT provides the most accurate results. However, for e-commerce Y, IndoBERT performs reasonably well, while logistic regression and random forest demonstrate potential in performing sentiment analysis on small- to medium-sized datasets. IndoBERT is more suitable for large datasets; the greater the number of data samples,

the more precise the sentiment analysis outcomes. The following section summarizes the classification results for e-commerce X.

**Table 5.** Classification Report for E-Commerce X

<b>Polarity</b>	<b>Model</b>	<b>Classification</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Rating-Based	Random Forest	Negative	0.59	0.56	0.58	203
		Neutral	0.48	0.39	0.43	204
		Positive	0.67	0.83	0.74	204
	SVM	Negative	0.64	0.61	0.62	203
		Neutral	0.49	0.53	0.51	204
		Positive	0.76	0.72	0.74	204
	Logistic Regression	Negative	0.61	0.57	0.59	203
		Neutral	0.46	0.52	0.49	204
		Positive	0.76	0.72	0.74	204
	IndoBERT	Negative	0.56	0.92	0.70	203
		Neutral	0.65	0.19	0.30	204
		Positive	0.77	0.83	0.80	204
VADER-Based	Random Forest	Negative	0.86	0.94	0.90	817
		Neutral	0.91	0.89	0.90	817
		Positive	0.98	0.90	0.94	817
	SVM	Negative	0.94	0.96	0.95	817
		Neutral	0.93	0.97	0.95	817
		Positive	0.99	0.92	0.95	817
	Logistic Regression	Negative	0.90	0.95	0.92	817
		Neutral	0.90	0.93	0.91	817
		Positive	0.98	0.90	0.94	817
	IndoBERT	Negative	0.96	0.98	0.97	817
		Neutral	0.96	0.94	0.95	817
		Positive	0.96	0.96	0.96	817

Source: research data

Based on Table 5, the sentiment classification report on E-Commerce X data shows a comparison of the performance of four machine learning models (Random Forest, SVM, Logistic Regression, and IndoBERT) applied after VADER-based feature extraction. Overall, the IndoBERT model provides the best performance across all polarity classes. IndoBERT achieves very high F1-Score values: 0.97 for the Negative class, 0.95 for Neutral, and 0.96 for Positive. This indicates the model's superior ability to balance Precision (accuracy of positive predictions) and Recall (ability to capture all positive cases) within each sentiment category, with respective values ranging from 0.96 to 0.98.

The SVM model also demonstrates strong performance, with a consistent F1-Score of 0.95 across all sentiment classes, making it the second-best choice. In contrast, Random Forest and Logistic Regression produce more moderate performance. Logistic Regression achieves the lowest but still acceptable F1-Score of 0.91 for the Neutral class and 0.92 for Negative. All models report the same Support value of 817 per class, indicating a balanced data distribution for Negative, Neutral, and Positive samples.

In conclusion, for the sentiment classification task on this dataset, the transformer-based IndoBERT architecture proves to be the most effective and reliable model.

**Table 6.** Classification Report for E-Commerce Y

<b>Polarity</b>	<b>Model</b>	<b>Classification</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>	<b>Support</b>
Rating-Based	Random Forest	Negative	0.60	0.65	0.62	203
		Neutral	0.59	0.46	0.52	204
		Positive	0.71	0.82	0.76	204
	SVM	Negative	0.63	0.67	0.65	203
		Neutral	0.59	0.52	0.56	204
		Positive	0.79	0.83	0.81	204
	Logistic Regression	Negative	0.64	0.70	0.67	203
		Neutral	0.60	0.51	0.55	204
		Positive	0.79	0.83	0.81	204
IndoBERT	Negative	0.52	0.87	0.65	203	
	Neutral	0.48	0.27	0.35	204	
	Positive	0.83	0.63	0.72	204	
VADER-Based	Random Forest	Negative	0.78	0.80	0.79	125
		Neutral	0.73	0.85	0.78	126
		Positive	0.95	0.77	0.85	126
	SVM	Negative	0.77	0.82	0.79	125
		Neutral	0.73	0.82	0.77	126
		Positive	0.94	0.77	0.85	126
	Logistic Regression	Negative	0.73	0.82	0.77	125
		Neutral	0.72	0.82	0.76	126
		Positive	0.97	0.72	0.83	126
IndoBERT	Negative	0.73	0.58	0.64	125	
	Neutral	0.59	0.89	0.71	126	
	Positive	0.89	0.63	0.73	126	

Source: research data

The sentiment classification report in Table 6, E-Commerce Y Data, presents a comparison of model performance in two feature extraction scenarios: Rating-Based and VADER-Based. Overall, the results indicate that the use of VADER-Based features consistently provides superior classification performance compared to Rating-Based features across all models and sentiment classes. Conversely, in the Rating-Based scenario, all models exhibit relatively low performance. Logistic Regression and SVM models perform best, achieving an F1-Score of 0.81 for the Positive class. However, IndoBERT in this scenario experiences a significant performance decrease, particularly in the Neutral class, with the lowest F1-Score of 0.35, driven by a very low Recall value (0.27). It is noteworthy that the amount of support (data samples) for VADER-Based features is much smaller (approximately 377 samples) compared to Rating-Based features (approximately 611 samples), which may contribute to the observed performance differences between scenarios.

Based on the test results, there is a pronounced performance disparity between deep learning models and conventional machine learning models, which is directly influenced by data volume. On the E-Commerce X dataset, which contains a substantial data volume of 4,084 samples per class (of which 817 per class constituted the test set used in classification reports), the IndoBERT model demonstrates clear superiority, achieving the highest accuracy of 0.96. This confirms that transformer architectures are highly effective in capturing complex linguistic contexts in e-commerce reviews when supported by large datasets. Conversely, on the E-Commerce Y dataset, which only includes 627 samples per label, the Random Forest model actually outperforms IndoBERT, achieving an accuracy of 0.81 compared to 0.75. This finding suggests that for limited-scale datasets, traditional ensemble algorithms tend to be more stable and less prone to overfitting compared to deep learning models that require large datasets for optimal generalization. Additionally, all tests consistently show that the VADER-based labeling method yields higher accuracy compared to the Rating-Based method, confirming the presence of

rating bias, where star numbers often do not accurately reflect the actual content of the review text.

The implementation of this sentiment analysis model provides practical utility for the management of E-Commerce X and Y by mapping priority issues based on authentic consumer voices that are not captured by star ratings. Based on the IndoBERT model's Wordcloud for E-Commerce X, positive sentiment is dominated by appreciation for features such as "shipping cost," "free," and "shopeepay." In the negative cluster, critical keywords such as "application," "error," and "courier" emerge. The recommendation for Platform X is to enhance application system stability and conduct periodic audits of logistics partner performance to ensure that massive transaction growth is accompanied by consistent delivery quality.

For E-Commerce Y, the Wordcloud results from the Random Forest model reveal more specific insights related to technical constraints. Negative sentiment prominently highlights terms such as "error," "application," "cancel," and "funds," indicating serious issues in the transaction flow and user fund returns. The strategic recommendation for Platform Y is to promptly repair the payment gateway infrastructure and order processing system to prevent further decline in traffic. By leveraging keyword extraction from this highest-accuracy model, both companies can make data-driven product improvements, prioritize addressing the most frequent user complaints, and ultimately restore customer trust and loyalty in the long term.

The accuracy achieved by IndoBERT (0.96 on E-Commerce X, VADER-based) surpasses results reported by Nabillah et al. (2023), who achieved 0.89 with BERT-base for Indonesian social media. Similarly, Akbar et al. (2024) reported 0.84 for BiLSTM-based e-commerce review analysis, indicating that fine-tuned IndoBERT with VADER-based labeling offers superior performance. For classical ML models, the Random Forest result of 0.94 (F1, VADER-based, E-Commerce X) aligns with Ananta et al. (2024), who reported 0.87 for Shopee review classification. The superior performance of Random Forest over IndoBERT on E-Commerce Y (0.81 vs 0.75) is consistent with findings by Rayhan et al. (2024), suggesting that ensemble methods generalize better on datasets with fewer than 1,000 samples per class, likely due to lower variance and reduced susceptibility to overfitting compared to transformer models requiring large corpora (Luo et al., 2024; A. Rayhan & Gross, 2023).

## Deployment

The deployment architecture comprises: (1) a Flask-based REST API backend for model serving; (2) pre-loaded trained models (Random Forest, SVM, Logistic Regression, and IndoBERT) serialized using joblib/pickle formats; and (3) a monitoring module that tracks prediction drift using rolling accuracy metrics. The system accepts JSON input via POST requests and returns sentiment probabilities and final classification labels. Model retraining is triggered automatically when rolling accuracy drops below 0.75 on a 500-sample sliding window.

The deployment stage is the final phase of the CRISP-DM process, focusing on applying the model results in a manner usable by users or researchers. At this stage, the model that has been tested and demonstrated the best performance is implemented to provide automatic sentiment analysis results on new review data. The main objective of this stage is to ensure that the model runs stably, produces consistent predictions, and can be integrated with other systems or platforms if needed for further analysis.

Furthermore, prediction results from the model whether derived from machine learning algorithms such as Random Forest, SVM, Logistic Regression, or from deep learning such as IndoBERT can be visualized as analytical reports or interactive dashboards. This allows e-commerce developers to understand consumer perceptions in real-time, identify areas for improvement, and make data-driven strategic decisions. The model can also be updated periodically with new data to remain relevant to the evolving dynamics of user reviews.

```

=====
HASIL ANALISIS SENTIMEN
=====
Teks asli: sangat bagus dan cepat
Teks bersih: sangat bagus dan cepat
-----

📊 TRADITIONAL ML MODELS:
-----

Random Forest:
Prediksi: Positive
Confidence: 68.00%
Probabilitas:
- Negative: 14.00%
- Neutral: 18.00%
- Positive: 68.00%

Logistic Regression:
Prediksi: Positive
Confidence: 52.14%
Probabilitas:
- Negative: 23.47%
- Neutral: 24.38%
- Positive: 52.14%

SVM:
Prediksi: Positive
Confidence: 6.47%
Probabilitas:
- Negative: 51.12%
- Neutral: 42.41%
- Positive: 6.47%

-----

🇮🇩 INDOBERT MODEL:
-----

Prediksi: Negative
Confidence: 41.24%
Probabilitas:
- Negative: 41.24%
- Neutral: 32.15%
- Positive: 26.61%

-----

📄 RINGKASAN
=====
Random Forest      → Positive
Logistic Regression → Positive
SVM                → Positive
IndoBERT           → Negative
=====

```

**Figure 5.** Visualization of Sentiment Analytical Report Using Rating-Based  
Source: research data

Based on Figure 5, the author provides a visualization for sentiment analytical reports based on ratings. In this simple program, users can input a sentence or multiple data samples. These inputs are then analyzed by four models: Random Forest, SVM, Logistic Regression, and IndoBERT. The probability results for each class are displayed, and the class with the highest probability is considered the final result. However, a drawback for small datasets lies in the IndoBERT algorithm. A minimum of 20 to 100 samples is required to obtain fairly accurate IndoBERT results, consistent with findings by Luo et al. (2024), which show that Transformer-based models require sufficient labeled examples for reliable fine-tuning. In our experiments, IndoBERT predictions with fewer than 20 samples exhibited near-random class probability distributions.

```

=====
ALL MODELS READY - START TESTING - VADER BASED
=====

⚠ IMPORTANT: Models are minimally trained with sample data.
For accurate predictions, train with 50-100 real labeled samples.

Masukkan teks untuk dianalisis (atau 'quit' untuk keluar): sangat bagus dan cepat

-----
📄 TRADITIONAL ML MODELS (Trained with VADER)
-----

Random Forest:
  Prediksi: Positive
  Confidence: 71.00%
  Probabilitas:
    - Negative: 29.00%
    - Neutral: 0.00%
    - Positive: 71.00%

Logistic Regression:
  Prediksi: Positive
  Confidence: 47.05%
  Probabilitas:
    - Negative: 36.69%
    - Neutral: 16.26%
    - Positive: 47.05%

SVM:
  Prediksi: Positive
  Confidence: 47.13%
  Probabilitas:
    - Negative: 36.01%
    - Neutral: 16.86%
    - Positive: 47.13%

-----
📄 INDOBERT (Deep Learning - Untrained)
-----

  Prediksi: Neutral
  Confidence: 36.17%
  Probabilitas:
    - Negative: 30.12%
    - Neutral: 36.17%
    - Positive: 33.71%

⚠ Model belum dilatih - hasil acak!

-----
📄 RINGKASAN
-----

Random Forest      → Positive
Logistic Regression → Positive
SVM                 → Positive
IndoBERT (Untrained) → Neutral

```

**Figure 6.** Visualization of Sentiment Analytical Report Using VADER-Based  
Source: research data

The Sentiment Analytical Report Visualization in Figure 6 shows the test results of four models Random Forest, Logistic Regression, SVM, and IndoBERT in the deployment stage using one positive text input, "very good and fast," with VADER-based feature extraction. Traditional machine learning models (Random Forest, Logistic Regression, and SVM) trained minimally successfully identified the sentiment as positive. Of the three, Random Forest showed the highest confidence at 71.00%, while Logistic Regression and SVM produced positive predictions with lower confidence levels (around 47%), where the negative probability was still relatively high (around 36%). These results simultaneously reinforce the important warning that models require further training with labeled data (50–100 samples) to achieve optimal accuracy. On the other hand, the IndoBERT (deep learning) model predicted neutral with very low confidence; however, this result is considered invalid because there is a clear note indicating that the model has not been trained and its predictions are regarded as random. Overall, under this minimal live testing condition, Random Forest proved most confident in predicting the correct positive sentiment.

Additionally, the deployment stage includes a continuous model performance monitoring and evaluation process. Monitoring ensures that the model does not experience a decrease in accuracy (model drift) due to changes in language patterns or user behavior. If a performance decrease is detected, the model will be updated, or a retraining process with a new dataset will be conducted. Thus, the deployment stage not only marks the conclusion of the research but also provides the foundation for the sustainable application of sentiment analysis systems on e-commerce platforms in the future.

### Research Implications for both *e-commerce* platforms

The research implications for both e-commerce platforms indicate that platform X can further optimize its advantages in promotions and rapid response to consumer needs, as data shows a dominance of positive sentiment related to price and shipping. Platform X can also maintain competitiveness by continuing to utilize discount strategies and attractive offers tailored to user preferences. Conversely, platform Y has the opportunity to improve aspects of customer service and product quality, since negative sentiment on this platform is often associated with post-purchase complaints. By enhancing product verification and accelerating complaint handling, platform Y can increase user perception and customer loyalty. Both platforms should leverage sentiment analysis results as a basis for data-driven decision making, enabling more adaptive and efficient business strategies. Furthermore, the sentiment analysis system can support: (1) predictive customer satisfaction scoring monitoring weekly sentiment trends to proactively identify service degradation before impacting retention; (2) automated complaint categorization routing negative reviews tagged with "error," "cancel," and "refund" keywords directly to relevant support teams; and (3) recommendation system optimization integrating positive sentiment signals into collaborative filtering algorithms to improve personalization (Wang et al., 2024).

### CONCLUSION

Based on the research that has been conducted, it can be concluded that the application of machine learning and deep learning algorithms successfully provides effective results in conducting sentiment analysis on user reviews of two e-commerce platforms. From a series of experiments, the IndoBERT model showed the best performance compared to classical algorithms such as Random Forest, SVM, and Logistic Regression. IndoBERT's superior performance is attributed to pre-training on large Indonesian corpora, enabling deep contextual understanding, as detailed in the Results section. The comparison between the two labeling methods, rating-based and VADER-based, shows that the VADER-based approach produces more accurate and consistent sentiment classification. Text-based labeling like VADER is able to capture the emotional nuances of users more naturally compared to numerical ratings, which tend to be subjective. This result is clearly visible from the increased accuracy and F1-score values produced by most models on the VADER-based dataset. Overall, this research successfully built an automatic Sentiment Analysis Comparison of Two E-Commerce Platforms Using Random Forest, Support Vector Machine, Logistic Regression, and IndoBERT system that can identify positive, negative, and neutral opinions from user reviews. This system can be applied as a decision-making tool for e-commerce managers, particularly in understanding customer satisfaction and perceptions of their services. With further development, this model can also be integrated into a real-time monitoring system to support strategies for improving service quality and user experience in the future.

Several shortcomings in this research were identified that demonstrate how data quantity affects sentiment results on an e-commerce platform. One of the main limitations of this study lies in the quantity and availability of unbalanced data between the two platforms studied. Model performance on e-commerce X was found to be much superior and more stable compared to e-commerce Y, which was technically caused by the significant difference in data volume resulting from undersampling. On the VADER-based e-commerce X dataset, the model was supported by 4,084 total samples per label (817 per class in the test set), whereas e-commerce Y only had 627 total samples per label (125–126 per class in the test set). The support values in Tables 6 and 7 reflect test set sizes only. The limited amount of data on e-commerce Y resulted in deep learning models such as IndoBERT not obtaining a wide enough variety of language patterns to perform optimal learning and generalization. Conversely, the larger data volume on e-commerce X allowed the model to recognize sentiment features more accurately, achieving a maximum accuracy of 0.96. This difference in data scale shows that the performance of Natural Language Processing (NLP) algorithms in this study is highly dependent on the adequacy of the number of samples, where smaller datasets tend to produce lower and less consistent performance in classifying user opinions.

### ACKNOWLEDGEMENT

The authors would like to express gratitude to all research participants and reviewers for their constructive feedback. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

### AUTHOR CONTRIBUTION STATEMENT

Author 1: Conceptualization, Methodology, Data Collection, Analysis, Writing - Original Draft. Author 2: Supervision, Review and Editing, Validation. All authors have read and agreed to the published version of the manuscript.

### REFERENCES

- Ait, B. H. K., Jarir, Z., & Elfirdoussi, S. (2024). Design Of A Machine Learning-Based Decision Support System For Product Scheduling On Non Identical Parallel Machines. *Engineering, Technology & Applied Science Research*, 14(5), 16317–16325. <https://doi.org/10.48084/Etasr.7934>
- Akbar, H., Aryani, D., Mohammed Al-Shammari, M. K., & Ulum, M. B. (2024). Sentiment Analysis For E-Commerce Product Reviews Based On Feature Fusion And Bidirectional Long Short-Term Memory. *Jurnal Teknik Informatika (JUTIF)*, 5(5), 1385–1391. <https://doi.org/10.52436/1.Jutif.2024.5.5.2675>
- Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining The Text Of Online Consumer Reviews To Analyze Brand Image And Brand Positioning. *Journal Of Retailing And Consumer Services*, 67, 102989. <https://doi.org/10.1016/J.Jretconser.2022.102989>
- Ananta, F. A., Id, H. A., & Kania, N. A. (2024). Klasifikasi Sentimen Pada Aplikasi Shopee Menggunakan Fitur Bag Of Word Dan Algoritma Random Forest. *Ranah Research : Journal Of Multidisciplinary Research And Development*, 6(5), 1678–1683. <https://doi.org/10.38035/Rrj.V6i5.994>
- Gallin, S., & Portes, A. (2024). Online Shopping: How Can Algorithm Performance Expectancy Enhance Impulse Buying? *Journal Of Retailing And Consumer Services*, 81, 103988. <https://doi.org/10.1016/J.Jretconser.2024.103988>
- Glori, N., & Widjaja, W. (2024). Analisis Perbandingan Kualitas Promosi, Keragaman Produk Dan Kemudahan Penggunaan Antara E-Commerce Shopee Dan Tokopedia. *Bina Ekonomi*, 28(1), 21–32. <https://doi.org/10.26593/Be.V28i1.6229.21-32>
- Hutto, C., & Gilbert, E. (2014). Vader: A Parsimonious Rule-Based Model For Sentiment Analysis Of Social Media Text. *Proceedings Of The International AAAI Conference On Web And Social Media*, 8(1), 216–225. <https://doi.org/10.1609/ICWSM.V8i1.14550>
- Ibrahim. (2023). The Future Of Marketing: Emerging Trends And Technologies. *Epra International Journal Of Research & Development (IJRD)*.
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal Of Electrical And Electronics Engineering*, 5(1), 32–35. <https://doi.org/10.37905/Jjee.V5i1.16830>
- Leon, M. (2025). Retracted: Sentiment Analysis: From Rule-Based Lexicons To Large Language Models. *Intelligent Systems With Applications*, 28, 200599. <https://doi.org/10.1016/J.Iswa.2025.200599>
- Liang, X., Guo, J., Sun, Y., & Liu, X. (2021). A Method Of Product Selection Based On Online Reviews. *Mobile Information Systems*, 2021, 1–16. <https://doi.org/10.1155/2021/9656315>
- Liu, Z. (2024). Transformations In Consumer Buying Behavior: Investigating How Online Shopping Platforms. *Advances In Economics, Management And Political Sciences*, 109(1), 181–186. <https://doi.org/10.54254/2754-1169/109/2024bj0135>
- Luo, X., Deng, Z., Yang, B., & Luo, M. Y. (2024). Pre-Trained Language Models In Medicine: A Survey. *Artificial Intelligence In Medicine*, 154, 102904. <https://doi.org/10.1016/J.Artmed.2024.102904>
- Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2023). Bert Base Model For Toxic Comment Analysis On Indonesian Social Media. *Procedia Computer Science*, 216, 714–721. <https://doi.org/10.1016/J.Procs.2022.12.188>

- Novirianto, I. F. (2023). *Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Mybluebird Dengan Implementasi N-Gram Dan Algoritma Logistic Regressio*. Fakultas Sains Dan Teknologi Uin Syarif Hidayatullah Jakarta.
- Novitasari, D., Maulana, R., Hastuti, H., & Puspitasari, N. (2024). *Seminar Nasional Amikom Surakarta (SEMNAS) 2024*.
- Rayhan, A., & Gross, D. (2023). The Rise Of Python: A Survey Of Recent Research. *License CC By*, 4.
- Rayhan, F. M., Wijoyo, S. H., & Putra, W. H. N. (2024). Analisis Sentimen Root Cause Analisis Kepuasan Pengguna Aplikasi Tokopedia Pada Ulasan Menggunakan Metode Random Forest. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 8(8).
- Sanjaya, T. P. R., Fauzi, A., & Masruriyah, A. F. N. (2023). Analisis Sentimen Ulasan Pada E-Commerce Shopee Menggunakan Algoritma Naive Bayes Dan Support Vector Machine. *Infotech: Jurnal Informatika & Teknologi*, 4(1), 16–26. <https://doi.org/10.37373/Infotech.V4i1.422>
- Santhosh, V., & Basavarajappa, B. (2022). A Study On Consumer Behaviour Towards Online Shopping. *International Journal Of Research And Analytical Reviews (IJRAR)*, 9(1), 864–874.
- Schulze, J. (2020). *Online And Offline Shopping: Decision Making Factors That Influence Consumer Purchases*.
- Wang, L., Pertheban, T. R. A. L., Li, T., & Zhao, L. (2024). Application Of Business Intelligence Based On Big Data In E-Commerce Data Evaluation. *Heliyon*, 10(21), E38768. <https://doi.org/10.1016/J.Heliyon.2024.E38768>