Topic Analysis Using LDA-LSTM on Shopee User Comment

[1]Shannon Dominique Saputra, [2]Albertus Dwiyoga Widiantoro
Information System Department, Computer Science Faculty
Soegijapranata Catholic University, Semarang, Indonesia
[1]dominiqueshannon24@gmail.com, [2]yoga@unika.ac.id

Abstract-Shopee's growth has also shaped online shopping in Indonesia. This study analyzed user reviews to measure satisfaction and identify key service issues using a hybrid framework: LDA for topic modeling and LSTM for sentiment classification. Class imbalance was addressed using a combination of Random Oversampling and Neighborhood Cleaning Rule (ROS-NCL).
The results showed that LSTM + ROS-NCL outperformed ROS, NCL, and SMOTE, with 95% accuracy and a precision, recall, and F1-score of 0.94 each. These findings demonstrate that oversampling combined with noise cleaning effectively improves performance on imbalanced data, while also providing practical insights for feature development, logistics improvements, and Shopee's promotional strategy.

Keywords- LDA, LSTM, NCL, ROS, Sentiment Analysis, Shopee, Text Classification

I. INTRODUCTION

The Covid-19 pandemic has been a catalyst for the unprecedented acceleration of e-commerce adoption in Indonesia. Before the pandemic, the number of e-commerce users in Indonesia was estimated at around 75 million people, but in 2020 this figure increased to 85 million active users [1]. This growth was accompanied by a surge in transaction value or Gross Merchandise Value (GMV) from US$23 billion in 2019 to US$35 billion in 2020, reflecting nearly 50% growth in just one year [2], [3]. According to the e-Conomy SEA 2023 report compiled by Google, Temasek, and Bain & Company, Indonesia's e-commerce GMV is projected to continue growing to US$83 billion by 2025, a significant increase from US$32 billion in 2020 [4]. Data from the U.S. International Trade Administration also shows that in 2022, the number of online transactions increased by 23%, transaction volume rose 27% year-on-year to 222.9 million, and there were 63 million new users [5]. This surge not only reflects an increase in the number of buyers but also a shift in consumer behaviour, with people increasingly relying on online platforms like Shopee for daily needs, from groceries to lifestyle products.

The rapid growth in online shopping activity directly impacts the volume of user comments and reviews, which serve as a rich source of information regarding product quality, delivery speed and accuracy, transaction security, and overall customer satisfaction.

Comments, which were once merely considered ordinary feedback, have now become a research topic in their own right due to their massive volume, diverse language, frequent use of abbreviations or non-standard terms, and often informal language. This makes manual analysis ineffective, yet these comments contain valuable information about service quality, delivery challenges, and customer satisfaction.

Analysing these comments is crucial as it can reveal dominant topics frequently discussed, measure public perception, and identify areas of service that need improvement [6]. One commonly used method for extracting themes from large-scale text data is topic modelling. Comparative studies show that Latent Dirichlet Allocation (LDA) has an advantage in producing more coherent and semantically relevant topics compared to Non-negative Matrix Factorization (NMF), which, although more computationally efficient, tends to produce less consistent topics with low interpretability [7], [8]. As a probabilistic generative model, LDA views documents as mixtures of topics and topics as word distributions, thereby preserving semantic relationships between words which a significant advantage when processing diverse Shopee reviews that often use informal language [9].

After topics are successfully identified, the next step is topic classification to determine the topic tendencies of each review. In this context, Long Short-Term Memory (LSTM) is widely chosen due to its ability to process text data sequentially and capture long-term dependencies that conventional models cannot handle. Lu et al.'s research [10] demonstrated that Bidirectional LSTM achieved 92.1% accuracy on e-commerce reviews, outperforming Support Vector Machine and Naive Bayes methods. Recent studies even show that combining LSTM with optimisation algorithms can improve sentiment classification accuracy and precision by up to 32% [11]. By combining LDA for topic modelling and LSTM for topic classification, this research aims to generate deeper insights into Shopee users' opinions, which can be utilised for service improvements, feature innovations, and data-driven business strategies.

II. LITERATURE REVIEW

A. Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is one of the most popular topic modelling methods in text processing. This method is used to discover latent thematic structures in large document collections using a probabilistic approach. In the context of user reviews of applications such as Shopee, LDA can be used to explore frequently discussed topics such as product quality, delivery, or customer service [12, 13, 14].

B. Random Over Sampling (ROS) and Neighborhood Cleaning Rule (NCL)

The problem of class imbalance in review data (e.g., positive reviews far outnumber negative ones) can affect the performance of classification models. Therefore, the Random Over Sampling (ROS) method is used to duplicate minority data, while the Neighbourhood Cleaning Rule (NCL) is used to clean ambiguous majority data. The combination of these two methods has been proven to improve model accuracy and reduce overfitting in artificially balanced data [14].

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) model designed to handle long-term dependencies in sequential data such as text. LSTM is often used in sentiment classification because of its ability to retain contextual information from previous sentences, which is important in determining the polarity of a user review [15]. Research shows that LSTM outperforms classical models such as Naive Bayes and SVM in text-based sentiment classification [16].

D. Integration of LDA and LSTM

Several studies have proposed integrating LDA and LSTM, where the topic extraction results from LDA are used as additional features for the LSTM model. This approach has resulted in improved accuracy in sentiment analysis, particularly for unstructured texts with diverse contexts such as Shopee reviews [17].

III. METHOD

This research was conducted through several important stages, starting from data collection (scraping), data cleaning and preparation, topic labelling using LDA, data imbalance handling using ROS, NCL, and SMOTE methods, to sentiment classification designed to produce a reliable model in understanding user opinions. The overall methodological flow can be seen in Figure 1.

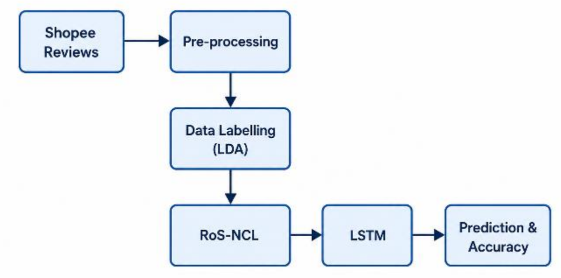

Figure.1. Research Methodology

A. Data Collection and Pre-Processing

This study utilises user review data from the Shopee application, collected from the Google Play Store. The data was gathered using web scraping techniques and resulted in over 7,000 review entries from January 2025 to May 2025.

A pre-processing stage was carried out to improve data quality by removing HTML characters, special symbols, irrelevant numbers, and converting all letters to lowercase. Non-standard words were corrected through normalisation, while tokenisation was used to break sentences into words. This stage also included stopword removal and initial labelling to facilitate topic modelling and classification.

B. Topic Modelling

After the text has been cleaned, the next step is topic modelling using the Latent Dirichlet Allocation (LDA) algorithm. The aim is to identify sets of themes or issues that frequently appear in user reviews.

LDA generates a number of topics, each consisting of keywords with specific weights. These keywords are analysed based on their frequency and occurrence in the documents. Thus, each review can be mapped to a specific topic distribution, which is then used as a feature in the classification process.

C. Labelling Based on Topic

The results of LDA are used to label the dominant topic in each document. This approach allows unstructured reviews to be incorporated into a more systematic thematic framework.

For example, if a comment has the highest probability of belonging to the topic ‘slow delivery’, then that comment will be labelled according to that topic. This information will be used in the next stage as part of the input features for the sentiment classification model.

D. Handling Imbalance Data

The distribution of sentiment labels in review data tends to be unbalanced, with positive reviews usually dominating. To address this, a combination of three techniques is used:

1. Random Over Sampling (ROS) is performed by doubling the data from the minority class to balance the number between classes.
2. Neighbourhood Cleaning Rule (NCL) functions as an undersampling technique that cleans majority data based on nearest neighbour analysis.
3. SMOTE (Synthetic Minority Oversampling Technique) is used to generate synthetic samples from minority data based on interpolation between original samples.

The combination of these three methods aims to improve the model's performance in recognising data from both classes fairly.

E. Sentiment Classification Using LSTM

Once the data is ready, the classification process is carried out using the Long Short-Term Memory (LSTM) algorithm, which is a neural network model for sequential data. LSTM was chosen because of its ability to capture context and patterns in text data more accurately than traditional models.

The model is trained to classify comments based on customised and standardised sentiment labels. The output of this model is a prediction of whether a comment is positive, negative, or neutral.

F. Model Evaluation

Model performance evaluation is carried out using the following matrix:

1. Accuracy

   Accuracy indicators are used to evaluate classification models. These indicators show how many predictions are correct compared to the total number of predictions available.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

   The precision calculation result is a measure of information that shows how accurate the information taken in the classification is and shows the accuracy of the correct prediction.

$$Precision = \frac{TP + TN}{TP + FP}$$

3. F-1 Value and F-1 Measurement
   The combination of precision and recall metrics that combines both by calculating the harmonic mean is called the F1-Score or F-measure.

$$F - measure = \frac{2xprecisionxrecall}{precision + recall}$$

4. Recall
   The ratio between the amount of data successfully recognised and the total amount of data that should be recognised.

$$Recall = \frac{TP}{TP + FP}$$

VI. RESULTS AND DISCUSSION

A. Data Collection and Pre-Processing

The data used in this study was obtained from user reviews on the Shopee app. The initial stage involved pre-processing the text to clean the data of special characters, stop words, and irrelevant word forms. This process was important to prepare the data for further modelling and to produce a more meaningful and consistent representation of the text.

B. Topic Modelling

The topic modelling process was carried out using the LDA algorithm with the best parameters obtained from the coherence score test. This model produced five main topics, each representing the dominant themes in Shopee user reviews. Topic 0 highlights service quality and customer satisfaction; Topic 1 relates to shopping experience and product quality; Topic 2 focuses on promotions and shipping costs; Topic 3 refers to financial services and payment methods; while Topic 4 describes logistics issues and product returns. These findings confirm that user experience is influenced by a combination of product quality, price, promotions, payment services, and logistics reliability, as shown in Table I.

C. Labelling Based on Topic

Once the topics have been identified, each comment is labelled according to the most relevant dominant topic. The labelling process is carried out by calculating the probability of correlation between a comment and a particular topic, as shown in Table II.

For example, the comment 'Awesome and reliable, very satisfied with shopping on Shopee' has a high probability of being related to Topic 0 (0.9106), while the comment 'Happy with the return process, though there were some minor issues' is associated with Topic 4 (0.4257). Topic 3 appears in comments such as "The app is great not only for shopping but also for payments" (0.4665) and "Suitable for e-money transactions, can easily buy and use ShopeePay" (0.6928), reflecting users' positive experiences with digital payment features like ShopeePay and e-money integration. Meanwhile, Topic 1 ('Shopee now has many products with varying quality,' 0.4943) indicates users' perceptions of the variety of products offered, both in terms of quantity and quality. Topic 2 is not dominant in these 10 comments, indicating that issues related to promotions and shipping costs may be less frequently discussed in this small sample. In this method, user reviews can be systematically classified based on the dominant discussion themes.

D. Handling Imbalance Data

The distribution of comments on each topic was uneven, so it was handled by combining Random Over Sampling (ROS) and Neighbourhood Cleaning Rule (NCL). ROS was used to increase the amount of data in the minority class, while NCL served to remove majority data that had the potential to be noise. The combination of these methods aimed to create a more representative and balanced dataset between classes.

E. Sentiment Classification Using LSTM

The next step is topic classification using Long Short-Term Memory (LSTM). This model was trained on data that had been balanced using ROS-NCL. The evaluation results show that this method is able to recognise each class well, with an overall accuracy of 95% and a macro average f1-score of 0.94. The performance of each class is also relatively consistent, for example, class 2 achieves an F1-score of 0.98, class 3 at 0.96, and classes 0, 1, and 4 at 0.93, 0.91, and 0.92 respectively, as shown in Table III below.

F. Model Evaluation

To test the effectiveness of the approach used, a comparison was made between the LSTM method and ROS-NCL, NCL alone, ROS alone, and SMOTE, as shown in Table IV. The results show that the LSTM method combined with ROS-NCL achieved the best performance (accuracy 95%, F1-score 0.94), significantly higher than NCL alone (accuracy 76%, F1-score 0.59), ROS alone (accuracy 36%, F1-score 0.33), or SMOTE (accuracy 34%, F1-score 0.28). This confirms that the ROS-NCL combination is superior because it can balance the data while reducing noise, resulting in more accurate and consistent topic classification.

Table 1 Table of Five Main Topics

Word Groups	Meaning Topics	Interpretation
Goods, shopping, suitable, hopefully, easy, fast, satisfied, delivery, service, order	Satisfaction to experience shopping	Users found the shopping process easy, the items were appropriate, the service was good, the delivery was fast, and they were generally satisfied with their orders.
Goods, buy, good, shopping, shop, online, help, products, new	Experience buy product by online	This topic describes the activity of purchasing goods/products through online stores. Words like *"good*," *"help*," and *"products"* indicate that users found the product or online shopping feature quite helpful.
Shipping, shopping, free, cheap, discount, often, online, increasingly, price, promotion	Promotions, discounts, prices cheap and free shipping	Users frequently discuss the benefits of online shopping, such as discounts, low prices, promotions, and free shipping. This indicates that price and promotions significantly influence shopping intentions.
Borrow, voucher, spaylater , account, pay, year, help, shopping, availability, once	Payments, vouchers, and Shopee PayLater features	Topics This related with facility payments, vouchers, accounts and services installments / loans like SPayLater. Word *help* show feature This considered helpful, but the presence of the words *borrow* and *pay* is also possible leading to the issue payment or loan.
Please, courier, cash on delivery, package, delivery, long, party, return, now, order	Shipping, courier, COD and return issues	This topic addresses complaints or requests related to packages, couriers, long shipping times, cash-on-delivery (COD) payments, and returns. Words like *"please* , " *"long* ," *"return* ," and *"delivery"* all point to less than satisfactory experiences.

Table 2 Labelling Results Table Based on Topic

ID Comment	Brief Comments	Dominant Topics	Probability
0	Cool and awesome, very satisfied shopping at Shopee	Topic 0	0.9106
1	Happy with the return process, despite a few minor issues	Topic 4	0.4257
2	Shopee now has many items of varying quality	Topic 1	0.4943
3	The app is excellent for shopping and can also be used for payments	Topic 3	0.4665

Table 3 Table Accuracy Results

Topic	Precision	Recall	F-1 Score	Support
Topic 0	0.95	0.91	0.93	283
Topic 1	0.89	0.92	0.91	105
Topic 2	0.98	0.97	0.98	219
Topic 3	0.93	1.00	0.96	229
Topic 4	0.94	0.89	0.92	92
Accuracy			0.95	928
Macro avg	0.94	0.94	0.94	928
Weighted avg	0.95	0.95	0.95	928

Table 4 Table Accuracy Results

Method	Accuracy	Recall	F1	Support
LSTM ROS-NCL	0.95	0.94	0.94	0.94
LSTM NCL	0.76	0.59	0.59	0.59
LSTM ROS	0.36	0.37	0.33	0.36
LSTM SMOTE	0.34	0.34	0.28	0.40

IV. CONCLUSION

This study examines Shopee user reviews in light of the rapid growth of Indonesian e-commerce and ShopeePay. By combining LDA for topic modeling and LSTM for sentiment classification, the study identified five key themes: product quality, shopping experience, promotions and shipping, financial services, and logistics. It also demonstrated variations in positive and negative sentiment within each theme, which can inform service improvements.

To address class imbalance, a hybrid balancing technique, ROS-NCL, was used and compared with ROS, NCL, and SMOTE. The best results were obtained with the combination of LSTM + ROS-NCL (accuracy 0.95; precision, recall, and F1-score 0.94), indicating that class upsampling combined with noise removal in the majority class is more effective than oversampling or undersampling alone.

REFERENCES

[1] ISEAS–Yusof Ishak Institute, “E-commerce in Indonesia: Impressive Growth but Facing Serious Challenges,” 2021. [Online]. Available: https://www.iseas.edu.sg/articles-commentaries/iseas-perspective/2021-102-e-commerce-in-indonesia-impressive-growth-but-facing-serious-challenges-by-siwage-dharma-negara-and-endang-sri-soesilowati/

[2] The Jakarta Post, “COVID-19 gives rise to new normal online shopping: Survey,” 2020. [Online]. Available: https://www.thejakartapost.com/news/2020/05/23/covid-19-gives-rise-new-normal-online-shopping-survey.html

[3] ISEAS–Yusof Ishak Institute, ibid.

[4] Google, Temasek, Bain & Company, “e-Conomy SEA 2023,” 2023. [Online]. Available: https://economysea.withgoogle.com

[5] U.S. Department of Commerce, International Trade Administration, “Indonesia – E-commerce Sector,” 2022. [Online]. Available: https://www.trade.gov/market-intelligence/indonesia-e-commerce-sectors

[6] H. H. Tran, “Sentiment Analysis in E-commerce: Methods and Applications,” Journal of Retailing and Consumer Services, vol. 72, p. 103115, 2023, doi:10.1016/j.jretconser.2023.103115.

[7] R. K. B. Putri, S. Purnama, and D. S. Nugroho, “Prediction of Research Trends Using LDA Based Topic Modeling,” IJACSA, vol. 12, no. 7, pp. 443–451,

2021, doi:10.14569/IJACSA.2021.0120751.

[8] A. A. Rahman and M. A. Hossain, “A Topic Modeling Comparison Between LDA and NMF,” Procedia Computer Science, vol. 214, pp. 865–874, 2022, doi:10.1016/j.procs.2022.02.143.

[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” JMLR, vol. 3, pp. 993–1022, 2003.

[10] Y. Lu, H. Liu, and F. Zhao, “Advanced Sentiment Analysis in Online Shopping: Implementing LSTM Models,” Electronic Commerce Research, 2025, doi:10.1007/s10660-025-09948-4.

[11] M. Khan, S. Hussain, and A. Bukhari, “Sentiment Analysis in E-commerce Platforms: A Review of Current Techniques and Future Directions,” arXiv preprint, arXiv:2505.03828, 2025.

[12] Hidayatullah, A. F., & Nugroho, E. (2023). Topic modeling for e-commerce customer reviews using LDA. International Journal of Intelligent Engineering and Systems, 16(2), 128–136. https://doi.org/10.22266/ijies2023.0430.12

[13] Hartono, P. C., Widiantoro, A. D., Harnadi, B., & Nugroho, E. W. (2025). Hybrid Approach of LDA, ROS-NCL, and CNN for Topic Mining in Fintech User Reviews. 2025 International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS), 1–6. https://doi.org/10.1109/ICADEIS65852.2025.10933096

[14] Widiantoro, A. D., Mustafid, M., & Sanjaya, R. (2023). Leveraging Latent Dirichlet Allocation for Feature Extraction in User Comments: Enhancements to User-Centered Design in Indonesian Financial Technology. Ingénierie Des Systèmes d Information, 28(5), 1423–1433. https://doi.org/10.18280/isi.280530

[15] Basu, S., et al. (2022). Handling class imbalance using ROS and Neighborhood Cleaning Rule for sentiment classification. arXiv preprint arXiv:2208.09619. https://arxiv.org/abs/2208.09619

[16] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv preprint arXiv:1412.3555.

[17] Akhtar, M. S., et al. (2020). Multi-task learning with deep bidirectional LSTM for sentiment analysis. PeerJ Computer Science, 6, e2542. https://peerj.com/articles/cs-2542

[18] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle. Proceedings of the ACL 2012 System Demonstrations, 115–120.