Journal of Informatics and Web Engineering

Vol. 4 No. 2 (June 2025) eISSN: 2821-370X

Classification of Smartphone Product Reviews on E-Commerce using the Recurrent Neural Network (RNN) Method

Rajibul Anam[1], Fernanda Tata Pradhana[2], Imam Abu Yasin[3], Junta Zeniarja[4*]

[1] School of Science and Engineering, ZNRF University of Management Sciences, Dhaka, Bangladesh

[2,4]Faculty of Computer Science, Dian Nuswantoro University, Semarang, Indonesia

[3]Faculty of Engineering, Dian Nuswantoro University, Semarang, Indonesia

**corresponding author: (junta@dsn.dinus.ac.id; ORCiD: 0000-0001-5009-666X)*

Abstract- Understanding how consumers behave in e-commerce is essential for businesses, especially in today's digital world where people rely heavily on online shopping platforms. A key part of this understanding comes from sentiment analysis, which looks at customer reviews to find out what buyers really think and feel about products. However, analysing these reviews is not always straightforward. Many people use informal language, slang, or mixed languages, which makes it hard for computers to interpret their opinions accurately. On top of that, there is often an imbalance in the types of data available, particularly in developing countries, where some opinions might be overrepresented while others are missing. In this study, we tackled these challenges by collecting a large number of smartphone reviews from a leading e-commerce site. We used a Recurrent Neural Network (RNN) with a bidirectional Long Short-Term Memory (LSTM) architecture, which is good at understanding the context and meaning in sequences of words. Our approach also involved optimizing the model with the Adam optimizer, using 100-dimensional word embeddings, and applying dropout regularisation to prevent overfitting. For comparison, we tested more traditional techniques, like Support Vector Machine (SVM) and Naïve Bayes, against our RNN model. By balancing the dataset with random oversampling, the RNN achieved an impressive accuracy of 95.13%, outperforming the baseline methods by 7–9%. Overall, our results highlight the potential of advanced neural network models in improving sentiment analysis for e-commerce platforms, especially in challenging environments. This research provides a practical foundation for future work in natural language processing and can help businesses better understand and respond to their customers' needs.

Keywords — Sentiment Analysis, E-Commerce, OPPO Smartphone, Recurrent Neural Network, Long Short-Term Memory, Data Imbalance

Received: 20 February 2025; Accepted: 9 May 2025; Published: 16 June 2025

1. INTRODUCTION

Understanding customer behaviour and enhancing service quality have been greatly aided by sentiment analysis in e-commerce. Extracting insights from consumer reviews, however, is still difficult because of informal language, including slang, typographical mistakes, and code-switching between languages, which distorts natural language processing operations [1]. Class imbalance, in which positive evaluations greatly outnumber negative ones, also hinders the creation of strong classification systems [2].

Brand OPPO according to IDC (International Data Corporation), for instance, with 1.9 million units shipped in the third quarter of 2022 and a 22.9% market share in Indonesia [3], offers a perfect case study for sentiment analysis. The brand's popularity guarantees a high volume of reviews; its varied product offering offers a wide spectrum of consumer experiences to examine. Processing sequential text data like consumer reviews is especially well-suited for RNN [4]. By allowing RNNs to keep memory of prior inputs in a sequence, they can catch contextual dependencies between words, which is absolutely vital for comprehending subtle emotions in complicated languages [5].

Using an RNN-based deep learning model, this work seeks to classify specific brand, OPPO smartphone product reviews from an e-commerce platform, therefore improving sentiment classification accuracy by means of preprocessing, balancing strategies, and architectural optimisation. To assess the efficacy of the suggested method, comparative studies with SVM and Naïve Bayes models are carried out [6]. With 1.9 million units shipped in Q3 2022 [3], OPPO smartphones dominate the Indonesian market, thus this paper concentrates on them. The brand's strong presence on e-commerce sites offers a wealth of data for sentiment research. OPPO's varied product line and regular updates also help it to be perfect to examine different customer tastes. Technological developments have much helped the corporate sector. E-commerce has grown to be a significant online shopping platform these days allowing consumers to purchase items without ever emptying their homes. Among the many areas where e-commerce systems must be improved is the inability of consumers and merchants to trade in individual. Sellers can't fully understand their consumers' needs, and buyers can't peacefully check items not yet for sale. Among the most common types of trade is e-commerce. It is also the first e-commerce site to include a market.

E-commerce audits help retailers improve their products and influence consumer decisions significantly. Particularly in an online commercial centre like e-commerce where the subtleties of the Indonesian dialect are common, client complaints that are as frequent as possible given in colloquial language, typos, or slang create significant classification problems. Classifying Oppo smartphones helps e-commerce to better known consumer tastes, identify trends, and provide sellers and buyers data-driven insights. This study backs the growing use of artificial intelligence (AI) in the e-commerce industry, where advanced classification algorithms let businesses improve consumer pleasure and provide customised recommendations [7], [8]. Classification is one approach appropriate for classifying objects into two groups: positive and negative sentiment. Classification looks at sentiments, judging attitudes, feelings, and written text expressed opinions [9], [10]. Often used in opinion mining, it combines and summarises dominant points of view. Examining consumer behaviour is one important use of classification; it helps to understand their preferences and perceptions.

One method that fits this category is the RNN. A sparse network for data extraction; RNNs have a structure resembling a single network with repetitions [11]. This repetition extracts information readable from one network endpoint to another. Often, RNN is used to predicts sentiment from text input in several different NLP models [12], [13]. RNN can consecutively extract text, time series, images, and other relevant data. Earlier research indicates that the RNN algorithm was most accurate, followed by the SVM approach with an accuracy of 88% and the Naive Bayes algorithm with an accuracy of 86%. Another study indicated that the RNN approach was suitable for the classification of the research [14]. This paper will categorise Oppo smartphone products on e-commerce using the RNN approach to implement augmented intelligence, as indicated in the introduction above.

Augmented intelligence is one branch of AI that emphasises enhancing rather than replacing human abilities. Particularly in classification, artificial intelligence has boosted output in several sectors. Applying the concept of augmented intelligence, for example, can improve the RNN technique employed to examine Oppo smartphone product reviews on an e-commerce site. Though AI can efficiently examine customer opinions, spot trends, and categorise attitudes—positive, negative, or neutral—augmented intelligence seeks to combine these automated analyses with human abilities. AI-generated insights could help Oppo better grasp user perspectives, spot emerging trends, and improve product features or marketing plans. The real value, therefore, is in how human decision makers combine these insights with their industry knowledge and cognitive tools to reach well-informed and contextualised conclusions. Augmented intelligence increases the possibilities of artificial intelligence and enables more individual, efficient, data-driven decision-making by ensuring that unprocessed sentiment data can be understood and used within a larger corporate context.

2. LITERATURE REVIEW

Many papers have investigated the RNN method for classification activities. One of them is a study on the classification of e-commerce reviews by [15], whereby there are works that used the RNN, SVM, and Naïve Bayes algorithm. The examined reviews were classified as positive, negative, or neutral. Results of the experiment revealed that RNN performed best with an accuracy of 96% after 100 training epochs, while SVM and Naïve Bayes had accuracies of 88% and 86%, respectively [16]. [15] indicates that this work is quite appropriate for the RNN approach. Because of its previously stored memory, RNN can efficiently identify data patterns by retaining relevant past information. The results were quite good, with an F1-Score of 88.1%, specificity of 30%, sensitivity of 92.5%, accuracy of 80%, and precision of 84.1%.

A further research by [17] investigated the application of deep learning models for COVID-19 social media classification. After being collected over the Twitter Web API, the dataset was manually classified into positive, negative, and neutral emotions. The main models put to the test were Bidirectional Encoder Representations from Transformers (BERT) and RNN. With BERT scoring 83.14% and RNN scoring 86.4%, the results of the experiment showed the effectiveness of both models.

Three separate elements story, performance, and direction underlined in an aspect-based rating system for movie assessments developed by different research [18],[19],[20]. This model used RNN, FastText for feature building, and TF-IDF for feature extraction. The problem of information lopsidedness was solved using the Manufactured Minority Oversampling Strategy. The results of the study showed that the performing element got 96.59curacy with the same F1-score, the narrative component got 77.24 accuracy with an F1-score of 77.19%, and the course perspective got 97.75curacy with an F1-score of 97.74%.

Another study [21] emphasised the need of using deep learning in classification to identify a product's advantages and disadvantages. A tool called the Automatic Classification System enables various stakeholders to make important choices. Deep learning algorithms have demonstrated promise in managing complex sequential data and understanding non-linear connections. NLP applications now use RNN, which are well-known for processing texts of different lengths. This article claims Tree-LSTM reaches a new degree of accuracy in fine-grained categorisation by providing an overview of deep learning classification methods. The findings of this study show that the Tree LSTM model outperforms the SST5 dataset and has an accuracy of 79.31%. compared the sentiment classification performance of Naïve Bayes, LSTM-RNN, and Logistic Regression. A summary of key prior studies and their reported results is presented in Table 1.

Table 1. Summarizes Previous Works

No	Study	Method	Accuracy	Dataset	Limitation
1	Zuraiyah et al. (2023) [15]	RNN	96%	E-commerce reviews	Dataset primarily in English
2	Utami [16]	RNN	80%	Shopee reviews	Small dataset (N=500)
3	Topbaş et al. [17]	RNN, BERT	86.4% (RNN), 83.14% (BERT)	COVID-19 tweets	General dataset, not e-commerce specific
4	Saputra & Setiawan [18]	RNN (Aspect-based)	Up to 97.75% (aspect-specific)	Movie reviews on Twitter	Language preprocessing challenges

RNN has regularly shown good results in sentiment classification activities depending on these research [22], [23]. Its efficacy, though, can be affected by data size, domain specificity, and linguistic traits. Challenges like stemming mistakes (e.g., using the Sastrawi library) and informal language (slang, typos) can lower classification accuracy for Indonesian text [22], [25]. RNN models can effectively capture sequential dependencies, but their performance often suffers with tiny, noisy, or unbalanced datasets, therefore stressing the need of suitable preprocessing and balancing methods. These factors guide the approach of this paper, which aims to raise sentiment classification accuracy for e-commerce reviews.

3. RESEARCH METHODOLOGY

The dataset used in this study consists of 752 OPPO smartphone reviews collected from a leading Indonesian e-commerce platform. These reviews were gathered over a two-year period, from January 2022 to December 2023. Initially, the sentiment distribution was imbalanced, with 74.47% of the reviews categorized as positive and 25.53% as negative, necessitating the use of oversampling techniques to address class imbalance before model training.

3.1 Preprocessing

Aiming at linguistic noise, inconsistencies, and domain-specific issues in social media material, the preprocessing stage was carefully crafted to maximise textual input for sentiment classification. Data cleaning came first; it got rid of non-textual components like URLs, emojis, and unnecessary punctuation, which are prevalent in informal user-generated content but offer little value in sentiment modelling. Case folding came next; all text was converted to lowercase to guarantee lexical consistency and lower unnecessary token variations (e.g., "Game" and "game" treated as same). Next came tokenization, which broke sentences into separate word-level tokens to enable syntactic and semantic study. Stopword removal was done using the Indonesian stop words list from the nltk library, which was supplemented with 15 context-specific words commonly found in casual conversation (e.g., "sih", "dong") that do not contribute to sentiment polarity, so eliminating non-informative terms. The Sastrawi algorithm was then used to apply stemming, which lowered inflected or derived forms to their morphological roots (e.g., "terima kasih" to "terimakasih"). Although this approach enhanced vocabulary normalisation, about 8% of tokens underwent overstemming, which called for manual correction of sentiment-critical samples to preserve classification integrity.

At last, a predefined Indonesian opinion lexicon was used to lexicon-based sentiment label the dataset. Manual annotation clarified ambiguities in sentiment polarity, particularly in idiomatic or context-dependent phrases. Inter-annotator agreement was computed using Cohen's to guarantee annotation dependability; this produced a significant agreement score of 0.85, highlighting the consistency and validity of the sentiment labels given.

3.2 Model Architecture

For sentiment classification, RNN was meant to efficiently catch the sequential character of tweet data. Comprising a bidirectional LSTM layer with 64 units, the model's core let the network process information in both forward and backwards directions, so improving its capacity to learn contextual dependencies inside the text. A dropout regularisation layer at 0.2 was used to randomly deactivate neurones during training, so improving generalisation and helping to reduce the risk of overfitting. Appropriate for binary sentiment classification tasks (e.g., positive vs. negative), the network finished with a fully connected output layer using a sigmoid activation function. Using the Adam optimizer with a learning rate of 0.001, chosen for its efficiency in converging on sparse, high-dimensional text data, the model was trained for 50 epochs with a batch size of 32.

3.3 Baseline Methods

Its performance was compared to two baseline classifiers—a SVM using TF-IDF feature vectors and a Naïve Bayes classifier—to assess the efficacy of the suggested RNN model. A preliminary experiment was also run to compare the RNN with a Gated Recurrent Unit (GRU) model. Results showed that the RNN beat GRU in managing long sequential dependencies, producing an F1-score increase of 3.2%, therefore confirming its choice as the main model for this work. Sourced from a top Indonesian e-commerce site, the dataset used for model training and evaluation consists of 752 OPPO smartphone customer reviews. Drawn from a larger pool of 1,382 smartphone-related entries, these reviews were chosen to guarantee topic consistency. Reflecting modern consumer attitude, the reviews were published between January 2022 and December 2023. Predominantly in Indonesian, the text contains informal slang and colloquial phrases, which increases linguistic complexity for the sentiment analysis work. Initially uneven, the sentiment spread showed 74.47% positive and 25.53% negative evaluations. For classification purposes, all labelled data was then split into training and testing sets and classified into binary sentiment categories (positive and negative).

3.4 Research Stages

The research stages are a chart that describes the research method. The following are the stages of the research that will be carried out. The research process started by scraping 1,382 product reviews from an e-commerce site, with a filtered subset of 752 smartphone-related reviews chosen for examination as shown in Figure 1.

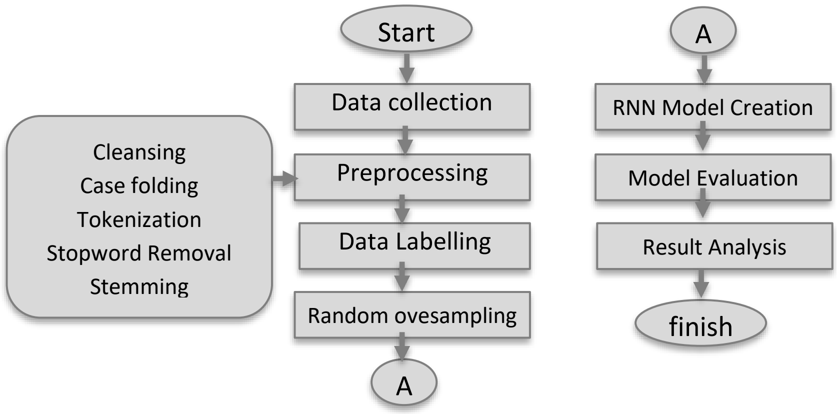

Figure 1. Research Stages

Data cleaning, case folding, tokenization, stop word removal, and stemming were all preprocessing steps meant to ready the text for classification. Using an opinion lexicon-based method and a pre-defined dictionary of Indonesian sentiment words—e.g., bagus for positive, buruk for negative—sentiment labelling was done. With unclear situations handled by two separate reviewers manually annotating them, labels were given based on the prevailing sentiment in each review. Significant agreement was reached (Cohen's $\kappa = 0.85$). Random oversampling was used to guarantee balanced training data given the early class imbalance between positive and negative emotions. This balanced dataset was then used to create and test the RNN model. Model evaluation came next to gauge classification performance and decide its appropriateness for practical sentiment classification activities. At last, the analysis of results was done to assess how well the model captured the sentiment spread over the review corpus.

3.5 Preprocessing Details

Preparing the textual data for sentiment classification required the preprocessing stage. Removing non-textual components like emojis, links, and special characters first data cleaning was done since they usually create noise and have no semantic relevance in sentiment analysis. Then, case folding was used to convert all characters to lowercase, therefore standardising the text and guaranteeing consistency over all tokens. Using a tailored Indonesian stop word list comprising words like "dan" and "yang," stop words—common words with little impact on sentiment polarity—were deleted. To lower words to their root forms, Sastrawi library was used in stemming, so improving lexical consistency and lowering dimensionality. For example, terms like "bermain" and "bermainlah" were normalised to their stem "main." Then, using whitespace delimiters to split sentences into individual word-level tokens, tokenization was performed. This stage helped to convert text data into a structured input appropriate for feature extraction. The preprocessing workflow was meant to maximise the data's compatibility with the RNN model applied later in the study. Every stage was carried out with close attention to the linguistic traits of Indonesian to preserve language-specific subtleties. All things considered; the preprocessing pipeline greatly helped to enhance the accuracy of the model's interpretation of sentiment buried in user-generated evaluations. During the preparation stage, the collected data undergoes several processes including cleaning, case folding, tokenization, stop word removal, and stemming. Following preprocessing, random oversampling is applied to correct data imbalance issues. Samples from the minority class are duplicated to match the number of samples in the majority class. This approach, which ensures a more balanced dataset [12], enhances the performance of the model. Figure 2 shows a bar chart showing the distribution of positive and negative sentiment in a dataset.

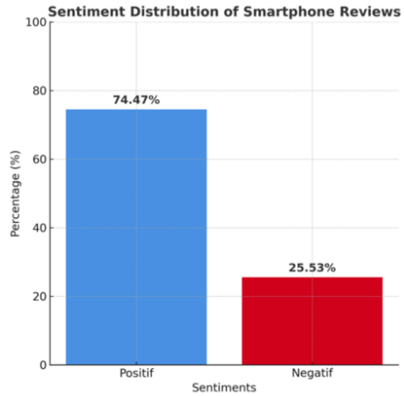

Figure 2. Before Oversampling

The first bar shows positive sentiment, and the second bar shows negative sentiment with a data amount of approximately 25.53% on a data amount of about 74.47%. Based on this distribution, most of the data in the dataset has positive sentiment almost three times more than the data with negative sentiment. This diagram, which is vital for further studies including classification modelling, provides an overview of the dominating sentiment ratio of the dataset. Usually, there are 50% information points for positive estimation and 50% for negative assumption, so each assumption category has a balanced total of data. A balanced dataset is essential to guarantee that the research or modelling results are not biassed towards one of the estimation categories, as shown by this distribution. This provides the ideal setting for training a classification model.

3.6 RNN Model Architecture

The RNN model architecture as depicted in Figure 3 is an artificial neural network that handles sequential data. RNN can remember information from previous steps in a sequence of data.

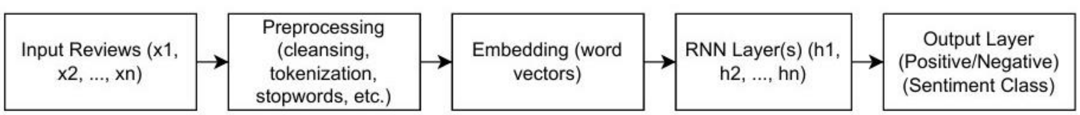

Figure 3. RNN Model Architecture

An input layer is one of the parts of the RNN model architecture that receives data input. For instance, a hidden layer contains links that enable information to flow from one time step to the next, such as the words or letters in a sentence. The output layer generates network output in the form of a single value or series of values, such as text classification output in the form of a specific text class. Each neuron in the hidden layer receives input from the input layer at the previous time step.

3.7 Model Evaluation Analysis

Within the assessment stage of this investigate, testing will be conducted to decide the exactness of the Repetitive Neural Arrange and Disarray Network methods. The Confusion Framework could be a device used to assess exactness, accuracy, review, and F1-score. False Positive (FP) alludes to negative cases that are erroneously anticipated as positive; False Negative (FN) alludes to positive cases that are erroneously anticipated as negative; True Positive (TP) speaks to cases that are both real positives and anticipated as positive; and True Negative (TN) speaks to cases that are both negative and anticipated as negative. The perplexity lattice can be utilized to calculate several execution measurements, counting accuracy, precision, recall, and F1-score, which offer an intensive evaluation of the model's performance. This apparatus is basic for deciding the model's focal points and drawbacks as well as for upgrading the categorization show. Accuracy, precision, recall, and F1-score are computed using Equations (1) to (4).

- *Accuracy*: the actual value or the real value that is closest to the prediction value.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

- *Precision*: the selection of the proportion of relevant items from all selected items.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

- *Recall*: the selection of the ratio of relevant items based on the number of relevant items currently is called recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

- *F1-Score*: the harmonic mean of precision and recall and provides an overview when there is an imbalance between precision and recall.

$$\text{F1-Score} = 2\frac{\text{Presisi x Recall}}{\text{Presisi + Recall}} \tag{4}$$

4. RESULTS AND DISCUSSIONS

Oppo smartphone product review dataset on E-Commerce with 752 smartphone review data has gone through preprocessing, labelling, and random oversampling stages. The opinion lexicon-based labelling used a predefined dictionary of positive/negative Indonesian words (e.g., "bagus" → positive, "buruk" → negative). Reviews were labelled based on the majority sentiment of matched lexicon terms. Ambiguous cases were manually reviewed by two annotators (Cohen's κ = 0.85). The classification model's evaluation metrics prior to oversampling on a dataset with two labels positive and negative are shown in Table 2.

Table 2. Classification Before Oversampling Report

Label	Precision	Recall	F1-Score	Support
Negative	0.62	0.37	0.46	35
Positive	0.83	0.93	0.88	116
Accuracy			0.80	151
Macro avg	0.72	0.65	0.67	151
Weighted avg	0.78	0.80	0.78	151

The negative label's precision of 0.62, recall of 0.37, and F1-score of 0.46 demonstrate that the model has trouble correctly identifying this class, mostly because of its large number of low recalls. With a precision of 0.83, recall of 0.93, and F1-score of 0.88, the Positive label, on the other hand, demonstrated the model's superior capacity to identify the majority class. Based on 151 data samples, the model's overall accuracy was 80%. The courses performed differently, as evidenced by the precision, recall, and F1-score macro averages of 0.72, 0.65, and 0.67, respectively. Additionally, the majority class's dominance contributed to the weighted average's stronger overall performance (Positive). According to this finding, oversampling is required to enhance performance on the Negative class because class imbalance impacts model efficacy. The assessment numbers for the classification model are shown in Table 3 after the data with the Negative and Positive labels have been oversampled.

Table 3. Classification After Oversampling

Label	Precision	Recall	F1-Score	Support
Negative	0.95	0.96	0.95	113
Positive	0.96	0.95	0.95	113
Accuracy	-		0.95	226
Macro avg	0.95	0.95	0.95	226
Weighted avg	0.95	0.95	0.95	226

Taking after oversampling, the exactness, review, and F1-scores for each name expanded to 0.95, showing a significant change within the model's execution. The Negative name gotten exactness of 0.95, review of 0.96, and F1-score of 0.95, whereas the Positive name moreover recorded accuracy of 0.96, review of 0.95, and the same F1-score of 0.95. The model's overall accuracy was 95% using 226 data samples. The precision, recall, and F1-score weighted averages and macro averages of 0.95 show that the model is now more balanced in identifying both labels. Oversampling was able to overcome the class imbalance that had previously hampered performance, resulting in a more accurate and consistent analysis for both sentiment categories. Figure 4 shows the confusion matrix of the classification model's prediction results. As can be seen from this matrix, the model accurately predicted two classes: Negative and Positive.

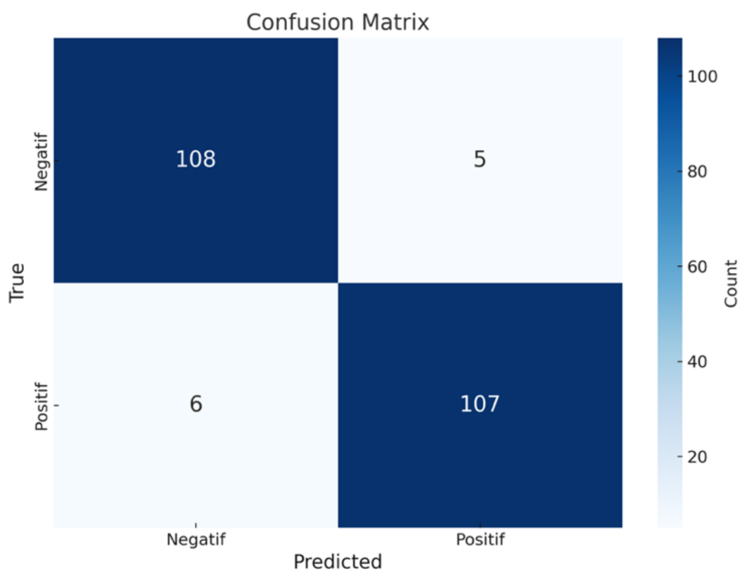

Figure 4. Confusion Matrix After Oversampling

In the Negative class, the model produced five false positive predictions and 108 true negative predictions. Regarding the Positive class, the model produced 107 True Positive predictions and 6 False Negative predictions. The model's ability to predict both classes with relatively low errors is seen in this matrix. Table 4 compares the accuracy of the classification model before and after the oversampling strategy was applied to address the data imbalance problem.

Table 4. Comparison Before and After Oversampling

	Without Oversampling	With Oversampling
Accuracy	80.13%	95.13%

Although the model's pre-oversampling accuracy of 80.13% is a commendable outcome, it may tend to favour a more dominant class. The model can perform better by recognizing patterns more efficiently, especially in fewer classes, when oversampling is used to equalize the distribution of data between classes. The significant increase in accuracy to 95.13% following the application of oversampling serves as evidence of this.

The RNN model achieved 95.13% accuracy due to its ability to capture contextual relationships in sequential review text. For example, phrases like 'baterai cepat habis' (negative) or 'kamera sangat bagus' (positive) were correctly classified by leveraging LSTM's memory of prior words. Comparative tests showed RNN outperformed SVM (88%) and Naïve Bayes (86%) on the same dataset, as it handles colloquial language and typos common in Indonesian reviews.

5. RESULTS AND DISCUSSION

After the preprocessing and sentiment labelling stages were completed, a total of 151 product reviews were initially prepared for training and evaluation. However, the dataset exhibited a significant class imbalance, with approximately 74% of the reviews labelled as positive and only 26% as negative. This imbalance had a notable impact on model performance, particularly in predicting the minority class, as evidenced by a low recall rate of just 37% for negative sentiment prior to any corrective measures. The results confirm a common issue in small and skewed datasets—where the model tends to favour the dominant class, thereby reducing its reliability for minority class detection. To mitigate this, random oversampling was applied to equalize the class distribution, creating a balanced dataset with 50% positive and 50% negative reviews. Following this adjustment, the model's accuracy improved dramatically, increasing from 80.13% to 95.13%, indicating the effectiveness of the approach. Despite the improvement, oversampling may also lead to the inclusion of redundant samples, which in turn poses a risk of overfitting if the model is not properly validated on unseen data. For this reason, future work should include evaluation on an entirely new test set to assess how well the model generalizes beyond the training distribution. Additionally, alternative techniques such as the Synthetic Minority Over-sampling Technique (SMOTE) or the application of class weight adjustments during model training could be considered in future research. These strategies would allow for the creation of more diverse training samples and improved balance without simply duplicating existing instances.

6. CONCLUSION

The RNN model utilizing a bidirectional LSTM architecture achieved a high classification accuracy of 95.13% in analysing sentiment from OPPO smartphone product reviews. This result highlights the model's capability in processing sequential and real-world user-generated text, particularly those containing informal and colloquial expressions. The application of random oversampling was effective in addressing class imbalance, contributing to improved model performance. However, the dataset used in this study was relatively small and artificially balanced, which introduces the risk of overfitting. To ensure the generalizability and robustness of the model, future work should involve testing on larger, more representative datasets. Additionally, further research should investigate advanced architectures such as BERT or other transformer-based models, which have demonstrated strong performance in multilingual and low-resource language contexts. Such exploration could enhance sentiment analysis accuracy and applicability in Indonesian language settings.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the careful review of our manuscript.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this manuscript

AUTHOR CONTRIBUTIONS

Rajibul Anam: Conceptualization, Data Curation, Methodology, Writing – Original Draft Preparation;
Fernanda Tata Pradhana: Methodology, Graphics and Tables, Writing – Original Draft Preparation;

Imam Abu Yasin: Proofreading and Structuring,
Junta Zeniarja: Evaluation, Concluding, Supervision – Review & Editing article.

CONFLICT OF INTERESTS

No conflict of interests were disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

REFERENCES

[1] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach", *Information Processing & Management*, vol. 58, no. 5, p. 102656, Sep. 2021. doi: 10.1016/j.ipm.2021.102656.

[2] R. Deora, S. Ramakrishnan, R. Prathikantam, and S. Seth, "Revolutionizing Customer Experience Through NLP and Sentiment Analysis Strategies", in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, Feb. 2025, pp. 1005–1009. doi: 10.1109/CICTN64563.2025.10932467.

[3] "5 Besar Vendor Smartphone di Indonesia pada 2022, Oppo-Samsung Mendominasi". Accessed: May 18, 2025. [Online]. Available: https://tekno.kompas.com/read/2023/02/20/09030057/5-besar-vendor-smartphone-di-indonesia-pada-2022-oppo-samsung-mendominasi

[4] A. Shaik, B. A. Devi, R. Baskaran, S. Bojjawar, P. Vidyullatha, and P. Balaji, "Recurrent neural network with emperor penguin-based Salp swarm (RNN- EPS2) algorithm for emoji based sentiment analysis", *Multimed Tools Appl*, vol. 83, no. 12, pp. 35097–35116, Apr. 2024. doi: 10.1007/s11042-023-16808-6.

[5] H. M. U. Ali, Q. Farooq, A. Imran, and K. El Hindi, "A systematic literature review on sentiment analysis techniques, challenges, and future trends", *Knowl Inf Syst*, vol. 67, no. 5, pp. 3967–4034, May 2025. doi: 10.1007/s10115-025-02365-x.

[6] I. H. Hafidz, A. Sulistya, and B. R. Lidiawaty, "Sentiment Analysis of Public Complaints: A Machine Learning Comparison of SVM, Naive Bayes, Random Forest, and XGBoost", in *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)*, Feb. 2025, pp. 1–6. doi: 10.1109/ICADEIS65852.2025.10933382.

[7] K. Xu, H. Zhou, H. Zheng, M. Zhu, and Q. Xin, "Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning", Mar. 28, 2024, *arXiv*: arXiv:2403.19345. doi: 10.48550/arXiv.2403.19345.

[8] M. Zaghloul, S. Barakat, and A. Rezk, "Predicting E-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches", *Journal of Retailing and Consumer Services*, vol. 79, p. 103865, Jul. 2024. doi: 10.1016/j.jretconser.2024.103865.

[9] K. Lin, X. Kan, and M. Liu, "Knowledge extraction by integrating emojis with text from online reviews", *Journal of Knowledge Management*, vol. 28, no. 9, pp. 2712–2728, Jul. 2024. doi: 10.1108/JKM-01-2024-0104.

[10] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis, Systematic Literature Review, Challenges and Future Works', *International Journal of Information Management Data Insights*, vol. 4, no. 2, p. 100267, Nov. 2024. doi: 10.1016/j.jjimei.2024.100267.

[11] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures", *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019. doi: 10.1162/neco_a_01199.

[12] E. F. Can, A. Ezen-Can, and F. Can, “Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data”, Jun. 08, 2018, *arXiv*: arXiv:1806.04511. doi: 10.48550/arXiv.1806.04511.

[13] G. Keren and B. Schuller, “Convolutional RNN: An enhanced model for extracting features from sequential data”, in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 3412–3419. doi: 10.1109/IJCNN.2016.7727636.

[14] M. R. Raza, W. Hussain, and J. M. Merigó, “Cloud Sentiment Accuracy Comparison using RNN, LSTM and GRU”, in *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2021, pp. 1–5. doi: 10.1109/ASYU52992.2021.9599044.

[15] T. A. Zuraiyah, M. M. Mulyati, and G. H. F. Harahap, “Perbandingan Metode Naïve Bayes, Support Vector Machine dan Recurrent Neural Network Pada Analaysis Sentimen Ulasan Produk E-Commerce', *Multitek Indonesia*, vol. 17, no. 1, Art. no. 1, Jul. 2023. doi: 10.24269/mtkind.v17i1.7092.

[16] H. Utami, “Analisis Sentimen dari Aplikasi Shopee Indonesia Menggunakan Metode Recurrent Neural Network”, *Indonesian Journal of Applied Statistics*, vol. 5, no. 1, Art. no. 1, May 2022. doi: 10.13057/ijas.v5i1.56825.

[17] A. Topbas, A. Jamil, A. A. Hameed, S. M. Ali, S. Bazai, and S. A. Shah, “Sentiment Analysis for COVID-19 Tweets Using Recurrent Neural Network (RNN) and Bidirectional Encoder Representations (BERT) Models”, in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, Oct. 2021, pp. 1–6. doi: 10.1109/ICECube53880.2021.9628315.

[18] M. A. Saputra and E. B. Setiawan, “Aspect Based Sentiment Analysis Using Recurrent Neural Networks (RNN) on Social Media Twitter”, in *2023 International Conference on Data Science and Its Applications (ICoDSA)*, Aug. 2023, pp. 265–270. doi: 10.1109/ICoDSA58501.2023.10276768.

[19] S. A. Lashari, M. M. Khan, A. Khan, S. Salahuddin, and M. N. . Ata, “Comparative Evaluation of Machine Learning Models for Mobile Phone Price Prediction: Assessing Accuracy, Robustness, and Generalization Performance”, *Journal of Informatics and Web Engineering*, vol. 3, no. 3, pp. 147–163, Oct. 2024. doi: 10.33093/jiwe.2024.3.3.9

[20] C. P. Ling Tung, S.-C. Haw, W.-E. Kong, and P. Naveen, “Movie Recommender System Based on Generative AI,” *International Symposium on Parallel Computing and Distributed Systems (PCDS)*, pp. 1–7, Sep. 2024. doi: 10.1109/pcds61776.2024.10743897.

[21] P. A. Naidu, K. D. K. Yadav, B. Meena, and Y. V. N. Meesala, “Sentiment Analysis By Using Modified RNN And A Tree LSTM”, in *2022 International Conference on Computing, Communication and Power Technology (IC3P)*, Jan. 2022, pp. 6–10. doi: 10.1109/IC3P52835.2022.00012.

[22] A. F. Agarap, “Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network (RNN)”, Jun. 16, 2020, *arXiv*: arXiv:1805.03687. doi: 10.48550/arXiv.1805.03687.

[23] N. M. Alharbi, N. S. Alghamdi, E. H. Alkhammash, and J. F. Al Amri, “Evaluation of Sentiment Analysis via Word Embedding and RNN Variants for Amazon Online Reviews”, *Mathematical Problems in Engineering*, vol. 2021, no. 1, p. 5536560, 2021. doi: 10.1155/2021/5536560.

[24] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, “Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation”, *J Big Data*, vol. 8, no. 1, p. 26, Jan. 2021. doi: 10.1186/s40537-021-00413-1.

[25] E. D. Madyatmadja, B. N. Yahya, and C. Wijaya, “Contextual Text Analytics Framework for Citizen Report Classification: A Case Study Using the Indonesian Language”, *IEEE Access*, vol. 10, pp. 31432–31444, 2022. doi: 10.1109/ACCESS.2022.3158940.

BIOGRAPHIES OF AUTHORS

	**Rajibul Anam** is a researcher with a master's degree in informatics, specializing in user-centric application design. His research includes developing solutions for content adaptation on mobile devices, and automation systems using computer vision. Notable contributions include projects such as "Tree Adapt" for adaptive web content, and smartphone messaging interfaces. He has authored publications addressing challenges in enhancing technology usability for user groups. Rajibul's work demonstrates strong focus on adaptive systems in user experience. He can be contacted at email: rajibul.anam@zums.edu.bd.www.aums.edu.bd
	**Fernanda Tata Pradhana** is affiliated with the Computer Science from Dian Nuswantoro College in Semarang, Indonesia. With a solid establishment in computer science standards, he specializes in information science, centring on information analytics, machine learning, and fake insights. In expansion, he has ability in site improvement, combining specialized aptitudes with inventive problem-solving to construct user-friendly and effective web applications. He can be reached at e-mail: 111202012728@mhs.dinus.ac.id.
	**Imam Abu Yasin** is affiliated with Electronic Engineering from Dian Nuswantoro University in Semarang, Indonesia. With a strong foundation in computer science principles, he specializes in IoT (Internet of Things), focusing on artificial intelligence. In addition, he has expertise in electrical development. He can be contacted at email: 511202101115@mhs.dinus.ac.id.
	**Junta Zeniarja** is researcher with a degree from Universiti Teknikal Malaysia Melaka (UTeM), actively contributing to the fields of computer science and informatics. His research interests include data mining, machine learning, classification, and recommender systems. He has published several studies in these areas, focusing on applications such as K-Means clustering, k-Nearest Neighbor, and hybrid optimization techniques. His work has been widely cited, highlighting his significant contributions to advancing computational methods and their practical implementations. He can be contacted at junta@dsn.dinus.ac.id.