Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

Analyzing Marketplace Reviews Using Word2Vec, CNN, and Deep K-Means with Sociolinguistic Approaches

Fahry[1], Titik Ceriyani Miswaty[*2], Harun[3]

[1]Computer Science Study Program, Universitas Bumigora, Indonesia
[2]English Literature Study Program, Universitas Bumigora, Indonesia
[3]Business Study Program, James Cook University, Singapore

Email: [1]titikceriyani@universitasbumigora.ac.id

Received : Sep 26, 2025; Revised : Oct 16, 2025; Accepted : Oct 16, 2025; Published : Dec 23, 2025

Abstract

This study investigates the effectiveness of deep learning methods in analyzing linguistically diverse customer reviews on Shopee to generate actionable product insights. By integrating Word2Vec, Convolutional Neural Networks (CNN), and Deep K-Means clustering, the proposed workflow moves beyond simple polarity detection toward aspect-based sentiment analysis. Customer reviews were preprocessed and represented using Word2Vec (skip-gram) to capture semantic proximity across informal registers, slang, abbreviations, and code-switching. A one-dimensional CNN then classified reviews into positive and negative sentiments, achieving 93–94% accuracy with balanced F1-scores across both classes. To extract aspect-level insights, reviews were projected into a latent space via an autoencoder and clustered using K-Means, with evaluation metrics (Silhouette ≈ 0.6; DBI ≈ 0.5) confirming adequate cohesion and separation. Positive clusters highlighted product design, durability, and ease of use, while negative clusters emphasized material quality, packaging, and delivery issues. These findings demonstrate that deep learning can adapt to sociolinguistic variation in Indonesian e-commerce discourse while providing structured, socially meaningful insights. This research is significant for the field of Informatics as it advances Natural Language Processing techniques for multilingual and code-switched data, addressing a key challenge in real-world text mining applications. The approach offers practical value for sellers in improving product quality, enhancing customer satisfaction, and refining marketing strategies.

Keywords : *Aspect-Based Sentiment Analysis, Convolutional Neural Network, Deep K-Means Clustering, Sociolinguistic Variation, Word2Vec Embedding*

1. INTRODUCTION

Extracting product strengths and weaknesses from online reviews holds high strategic value for business actors in digital marketplaces. On platforms such as Shopee, customer reviews represent one of the most influential sources of consumer insight, shaping not only individual purchasing decisions but also broader market competition [1], [2]. In a business environment characterized by intense competition, sellers are required to continuously innovate, align their products with market demands, and deliver offerings that combine affordability with superior quality and a user-friendly experience [3], [4]. Systematically identifying strengths, such as attractive design, durability, ease of use, or competitive pricing, allows sellers to effectively highlight these features in their marketing strategies, promotional content, and product descriptions [5]. Conversely, detecting weaknesses such as poor material quality, inaccurate sizing, or slow delivery enables more precise product evaluation [6] and targeted improvement.

Moreover, the systematic extraction of product strengths and weaknesses from online reviews helps build a more sustainable competitive advantage [7]. By leveraging natural language processing

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

and deep learning methods, businesses can efficiently analyze large volumes of unstructured customer feedback, uncovering hidden patterns that may not be visible through traditional market research. This analytical approach not only enhances decision-making at the operational level but also supports strategic planning by revealing consumer preferences, emerging trends, and potential market gaps. In this way, customer reviews are transformed from simple feedback into a valuable knowledge resource that guides product innovation, strengthens brand positioning, and fosters stronger relationships between sellers and consumers in the digital marketplace.

Despite their value, online reviews are typically abundant, unstructured, and highly diverse in language. Sellers often skim them as general satisfaction indicators without deeper analysis, leaving important insights overlooked [8]. This results in product revisions that are reactive or assumption-based rather than data-driven. Furthermore, the ability to monitor product performance over time, by comparing reviews before and after improvements or benchmarking against competitors, remains limited without systematic extraction techniques [9]. Sentiment analysis and natural language processing, therefore, hold significant promise in transforming raw textual feedback into actionable knowledge [10], [11]. Previous studies on sentiment analysis have made substantial progress in applying machine learning and deep learning models such as LSTM, CNN, or hybrid architectures for e-commerce review classification [12], [13]. These approaches successfully identify general sentiment polarity (positive, negative, neutral) and improve accuracy in various marketplaces and languages [14], [15]. In addition, recent works have employed aspect-based sentiment analysis (ABSA) to link sentiments with product features, offering more detailed insights [16], [17].

However, existing frameworks still face limitations when applied to linguistically diverse contexts such as Indonesian online marketplaces. Most models rely on standardized datasets and overlook sociolinguistic variability—such as slang, code-switching, and informal expressions—that dominate real customer reviews [18], [19]. As a result, current systems often misinterpret meaning, reduce classification accuracy, and fail to capture aspect-level sentiments tied to specific product attributes. Furthermore, many studies focus only on sentiment polarity without integrating clustering methods to reveal latent patterns across linguistically varied data. These limitations represent a significant research gap in adapting computational models to the linguistic and cultural realities of Indonesian e-commerce communication.

To fill this gap, the present study proposes an integrated framework that combines deep learning techniques—Word2Vec, Convolutional Neural Networks (CNN), and Deep K-Means clustering—with a sociolinguistic lens. This approach enables more accurate sentiment classification and aspect extraction by taking into account linguistic variation and context. The incorporation of sociolinguistic awareness enables the model to interpret informal, mixed-language, and culturally loaded expressions, such as “mantul banget” (slang for “very good”) or “ngaret dikit tapi oke,” which conventional models might misinterpret. By aligning computational analysis with sociolinguistic realities, this research introduces a novel hybrid model that generates socially meaningful insights and enhances the interpretability of marketplace review analytics. Based on these considerations, the research problems are formulated as follows. First, how can deep learning methods—specifically Word2Vec, Convolutional Neural Networks (CNN), and Deep K-Means clustering—be applied to extract information on product strengths and weaknesses from customer reviews in the Shopee marketplace, while accounting for sociolinguistic variation in digital language use? Second, to what extent are these deep learning methods effective in clustering and analyzing linguistically diverse customer reviews on Shopee to obtain relevant and socially meaningful insights into products?

Accordingly, the objectives of this study are twofold. The first objective is to develop and evaluate a deep learning model that can handle sociolinguistic variation in Indonesian e-commerce reviews while ensuring reliable sentiment classification. The second is to evaluate the effectiveness of deep learning

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

methods in clustering and analyzing linguistically diverse customer reviews on Shopee to generate relevant and socially meaningful product insights. Through this design, the research seeks to contribute to the methodological development of sentiment analysis in linguistically diverse contexts and to enrich the understanding of sociolinguistic dynamics in digital marketplaces. By acknowledging language variation in online reviews, this study generates more comprehensive and contextually accurate insights, providing practical value for sellers in product development, marketing strategy, and reputation management.

2. METHOD

This study is an experimental research aimed at identifying and achieving the best performance with the highest accuracy in analysing product reviews on the Shopee marketplace. It is descriptive in nature, as it details each stage of the review analysis process. A quantitative approach is applied, in which the results of the implemented methods are presented in numerical form to reflect the performance of the analytical models used [20],[21]. The process begins with determining the research type and approach, followed by collecting product review data through web scraping. it then proceeds to the pre-processing steps, including cleaning, sentiment labelling, and text normalisation [22]. The data are then represented using Word2Vec and analyzed through two stages: sentiment classification using CNN and product feature clustering with deep k-means clustering [23]. Classification performance is evaluated using a confusion matrix. At the same time, clustering quality is assessed with the silhouette score and the Davies–Bouldin index (DBI), and subsequently visualised through dimensionality reduction techniques to depict the features extracted from user reviews. From a sociolinguistic perspective, this visualisation is particularly valuable as it not only highlights sentiment polarity but also reveals linguistic variation across reviews, such as the use of slang, dialects, code-switching, or informal digital expressions. Incorporating this sociolinguistic approach enriches the interpretation of clusters by linking computational patterns with social meaning [24], ensuring that the analysis of product strengths and weaknesses remains both technically accurate and reflective of the diverse consumer voices in Indonesia's digital marketplace [25].

2.1 Data Collection

The main objective of data collection is to construct a new dataset that is both representative and of high quality, ensuring that the subsequent analytical models are trained on reliable and contextually relevant information [26]. At the initial stage, publicly available datasets from Kaggle were utilized to accelerate model prototyping and testing, as these datasets provided a benchmark for validating preprocessing pipelines and establishing baseline performance. Building on this foundation, a total of 25,000 Shopee product reviews were then collected directly from the official marketplace (shopee.co.id) through automated web scraping. The scraping process systematically retrieved essential attributes, including product titles, descriptions, categories, ratings, and user-generated reviews, thereby enabling a comprehensive view of both product characteristics and consumer feedback [27]. To further enhance data quality, the raw scraped data underwent cleaning procedures, including the removal of duplicates, normalization of textual formats, and filtering of irrelevant entries. Manual labeling was subsequently applied to a subset of 5,000 reviews to ensure accuracy in sentiment annotation and to provide ground truth for training supervised models. This multi-stage data collection process not only guaranteed the reliability of the dataset but also ensured its representativeness of actual consumer discourse in the Indonesian digital marketplace, making it highly valuable for developing robust extraction and sentiment analysis models [28].

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

2.2 Data Pre-processing

Data pre-processing aims to transform raw reviews into a clean, uniform, and analysis-ready format that effectively supports subsequent computational modeling. This process consists of three main stages. First, automated data cleaning was conducted to handle the large data volume by removing duplicates, correcting spelling errors, and eliminating inconsistencies, such as incomplete reviews or irrelevant symbols, to ensure that only high-quality entries were retained. Second, data labeling was performed manually by Shopee users and sellers, who categorized each review as positive, negative, or irrelevant. For analytical purposes, only positive and negative labels were retained to provide a clear distinction in polarity for model training and evaluation. To ensure consistency, detailed annotation guidelines were provided, and inter-annotator agreement was regularly checked to minimize subjectivity. Third, text normalization and tokenization were conducted through several sub-steps: case folding to standardize capitalization, tokenization to segment text into words, removal of punctuation and non-alphanumeric characters, normalization of informal language and abbreviations commonly used in digital discourse, stopword removal to eliminate non-informative words, and stemming to reduce words to their root forms [29], [30]. Furthermore, this study tested four pre-processing configurations by varying the inclusion or exclusion of specific operations, such as stopword removal or stemming, to evaluate their impact on model performance [31], [32]. All pre-processing tasks were implemented in Python using NLTK, Sastrawi, and scikit-learn libraries, and the processed datasets were stored in a structured database to support efficient retrieval and experimentation.

2.3 Data Representation

The review employed the Word2Vec skip-gram model, which is capable of capturing word meanings based on surrounding context, as shown in Figure 1. This model was applied to all four pre-processing scenarios to generate low-dimensional word vectors. Several key parameters were tested to obtain optimal results, including size, which determines the vector dimension [33]. Smaller sizes accelerate training, while larger sizes preserve more information. Window Size: Determines the extent to which the training process considers contextual words. Efficiency Method: Utilises hierarchical SoftMax or negative sampling to reduce training complexity. Negative: The number of random negative words used in negative sampling to support the learning of unrelated words. Experiments were conducted by configuring combinations of these parameters to produce the most effective representations for sentiment analysis.

The word embeddings with skip-gram and negative sampling is shown as (hs = 0). For each target–context pair $(w_i, w_o)$and Knegatives $w_k \sim P_n$, we minimize the logistic loss.

$$\ell_{\mathrm{NS}}(w_i, w_o) = -\log\sigma\left(\mathbf{u}_{w_o}^{\top}\mathbf{v}_{w_i}\right) - \sum_{k=1}^{K}\log\sigma\left(-\mathbf{u}_{w_k}^{\top}\mathbf{v}_{w_i}\right), \qquad (1)$$

And sum it over all target–context pairs within a symmetric window c:

$$\mathcal{L}_{\mathrm{NS}} = \sum_{t=1}^{T}\sum_{\substack{-c\le j\le c\\ j\neq 0}}\ell_{\mathrm{NS}}(w_t, w_{t+j}). \qquad (2)$$

Here v, u are input/output embeddings and $\sigma$the sigmoid. We tune size $\in \{50,100,150\}$, window $\in \{2,3,5\}$, and negative $\in \{5,10,20\}$; the selected setting is size=100, window=3, negative=20 (hs=0).

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

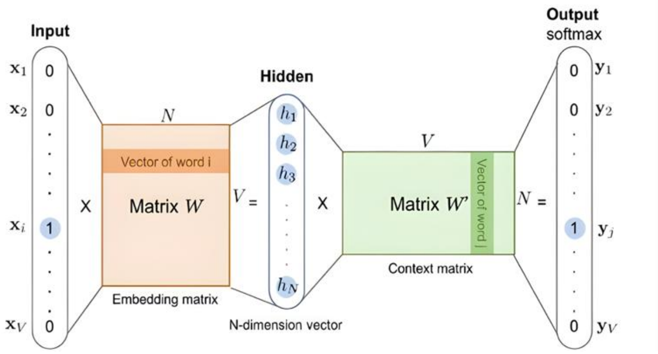

Figure 1. Skip-Gram model

2.4 Sentiment Classification

After the word representations were transformed into numerical vectors using Word2Vec, the classification process was carried out with a one-dimensional Convolutional Neural Network (1D CNN). The word vectors were fed into the input layer in the form of a matrix, where each row represented a word [34]. The 1D CNN was chosen because it is effective in capturing local features in text, such as key words or phrases, as explained in Kim's study [35]. It was also employed for its ability to extract important features that are highly relevant to sentiment [36]. To obtain the best results, the 1D CNN model was tested with several variations of key parameters, including Kernel Size, which defines the size of the filter in the convolutional layer to capture local patterns in the text. Dropout: Randomly deactivates a portion of neurons during training to prevent overfitting and improve generalisation. Learning Rate: Determines the step size when updating model weights. An appropriate value is essential for stable and efficient training. Each parameter combination was evaluated to determine its impact on the accuracy of sentiment classification. The architecture of the 1D CNN is shown in Figure 2.

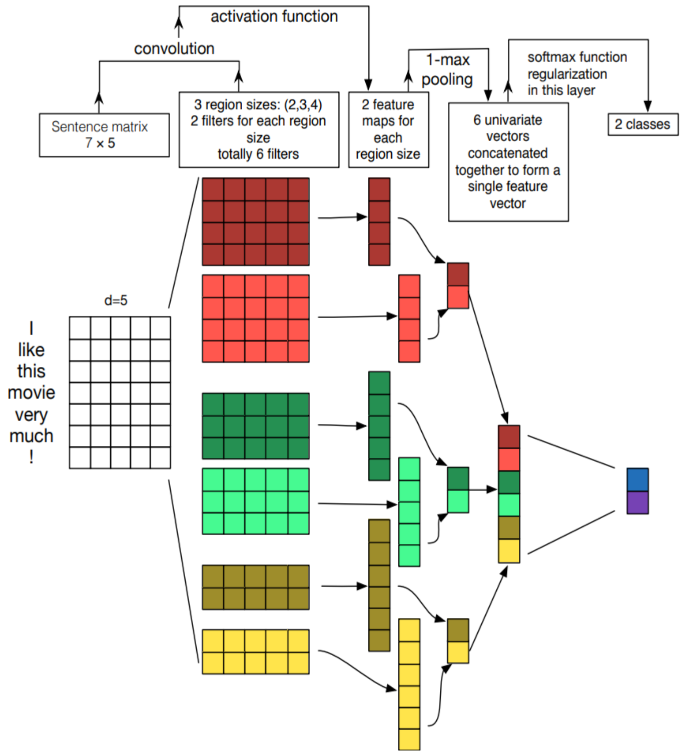

Figure 2: CNN 1D

Each review is encoded as an embedding matrix $\mathbf{X} \in \mathbb{R}^{L \times d}$. A 1-D convolution with kernel $\mathbf{W}^{(k)} = \{W_{p,q}^{(k)}\} \in \mathbb{R}^{m \times d}$ bias $b^{(k)}$produces feature maps

$$c_i^{(k)} = \mathrm{ReLU}\left(\sum_{p=0}^{m-1} \sum_{q=1}^{d} W_{p,q}^{(k)} X_{i+p,q} + b^{(k)}\right), s^{(k)} = \max_i c_i^{(k)}. \tag{3}$$

The pooled features $\mathbf{s} = [s^{(1)}; \dots; s^{(K)}]$are fed to a softmax classifier:

$$\hat{\mathbf{y}} = \mathrm{softmax}(\mathbf{W}_s \mathbf{s} + \mathbf{b}_s), \qquad \mathcal{L}_{\text{cls}} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{C} w_j \, y_{n,j} \log \hat{y}_{n,j}. \tag{4}$$

2.5 Clustering Evaluation

The quality of clusters produced by the Deep K-Means Clustering Neural Network was evaluated using two metrics: the Silhouette Score, which measures how similar data points are within a cluster compared to other clusters (with values closer to 1 indicating better clustering), and the Davies-Bouldin Index (DBI), which assesses the similarity between clusters [37] (lower values indicate better clustering). This evaluation is essential to ensure that the model can recognise the natural structure of the review data.

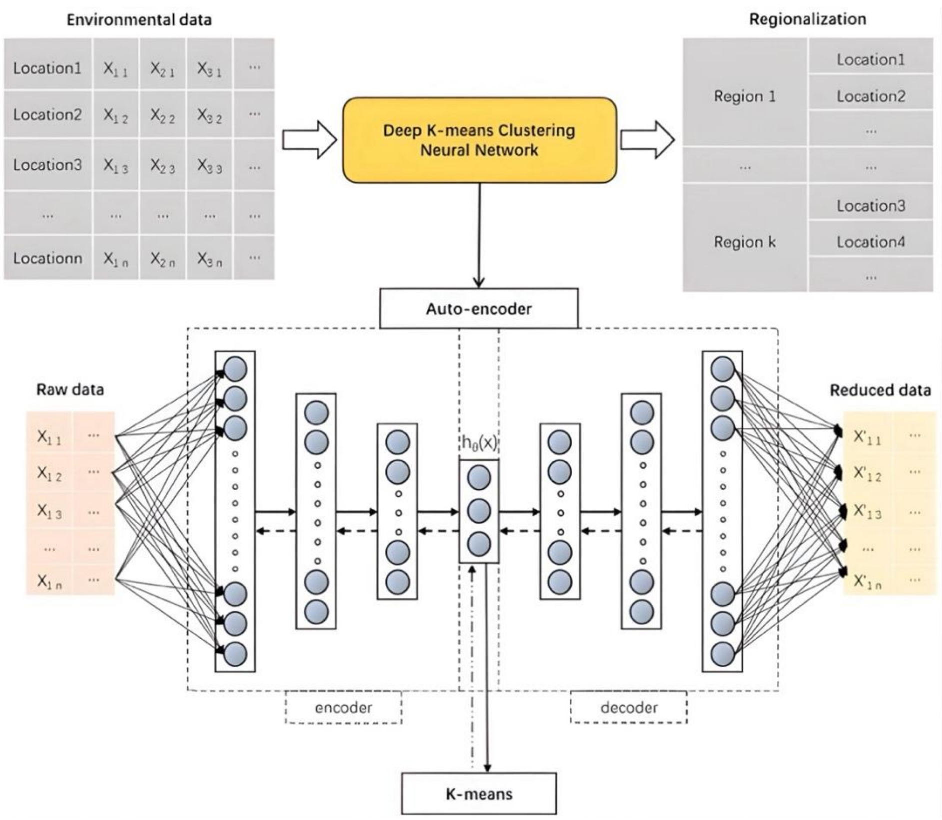

Figure 3: Deep K-Means Clustering Neural Network

2.6 Result Visualization

To better understand the distribution and quality of clusters, latent representations from the autoencoder were visualised in two dimensions using dimensionality reduction techniques such as t-SNE and PCA [38], [39]. These methods reduced the high-dimensional feature space into a form that is interpretable by humans while preserving the relative distances between data points as much as possible. Each cluster was represented with a distinct colour, allowing for clear differentiation and facilitating the

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

observation of patterns, overlaps, and potential outliers among clusters [40]. Through this visualization, it became possible to assess not only the compactness and separability of clusters but also the degree of coherence in grouping reviews with similar semantic and sentiment characteristics. Furthermore, visual inspection provided complementary evidence to quantitative metrics such as silhouette scores or Davies–Bouldin indices, thereby enhancing the overall evaluation of clustering quality. This combination of dimensionality reduction and visualization served as an essential diagnostic tool for validating the effectiveness of the clustering model and identifying areas for potential refinement.

3. RESULT

This study addresses the research problem by integrating three deep learning components: word2vec, convolutional neural networks (cnn), and deep k-means clustering neural networks to transform shopee customer reviews into operational aspect-sentiment information while accommodating sociolinguistic variation in digital language use. after optimal pre-processing is applied, the text is represented with word2vec (skip-gram) to capture semantic proximity between words across diverse registers, slang, abbreviations, and code-switching commonly found in e-commerce discourse. These representations are then classified by a one-dimensional CNN to reliably determine the polarity of each review, effectively identifying whether the sentiment is positive or negative despite linguistic variability. In the subsequent stage, deep k-means is employed: documents are projected into a latent space through an autoencoder and clustered separately for positive and negative corpora, generating aspect-based groupings of product features. From each polarity-cluster combination, dominant terms or phrases are extracted (e.g., using tf-idf) and interpreted as strengths when they prevail in positive reviews and as weaknesses when they dominate in negative reviews. By combining computational modelling with sensitivity to sociolinguistic variation, this workflow demonstrates how deep learning methods can move beyond simple polarity detection to provide structured, actionable insights into product strengths and weaknesses in the Shopee marketplace.

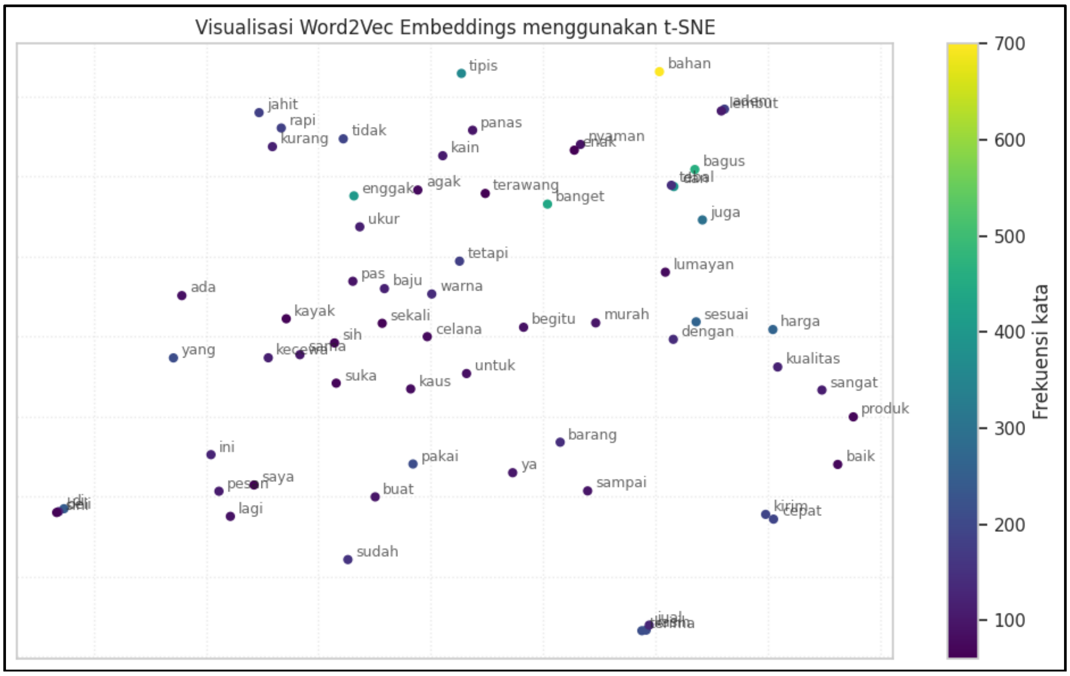

Figure 4. Word2Vec visualisation

The initial stage of the research begins with word representation using skip-gram-based Word2Vec. This representation is then visualized with t-SNE to illustrate the semantic proximity

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

between words. The results are presented in Figure 4, which displays the distribution pattern of words in the vector space and serves as the basis for the subsequent stage of analysis.

The Word2Vec-based 1D CNN model in the final experiment achieved an accuracy of 0.94 and a macro F1-score of approximately 0.94 on the test set (20% of the total data). The balanced performance between the negative and positive classes, as shown in Table 1, indicates the reliability of polarity labels for subsequent aspect-level analysis. Figure 2 presents the confusion matrix, while Figure 3 displays the ROC and Precision–Recall curves, which further confirm the consistency of the model's performance.

After sentiment labelling was obtained, aspect analysis was conducted using Deep K-Means. An autoencoder projected the documents into a lower-dimensional latent space, and these latent vectors were then clustered separately for the positive and negative corpora using K-Means, resulting in aspect clusters. At the aspect clustering stage in the autoencoder's latent space, the number of clusters was determined based on the Silhouette Score and Davies–Bouldin Index (DBI). The results yielded two clusters for the negative corpus, with a Silhouette score of 0.576 and a DBI of 0.577, and two clusters for the positive corpus, with a Silhouette score of 0.597 and a DBI of 0.535. Table 2 presents the Silhouette and DBI values for each class, while Figure 4 illustrates the distribution of clusters in both negative and positive sentiments.

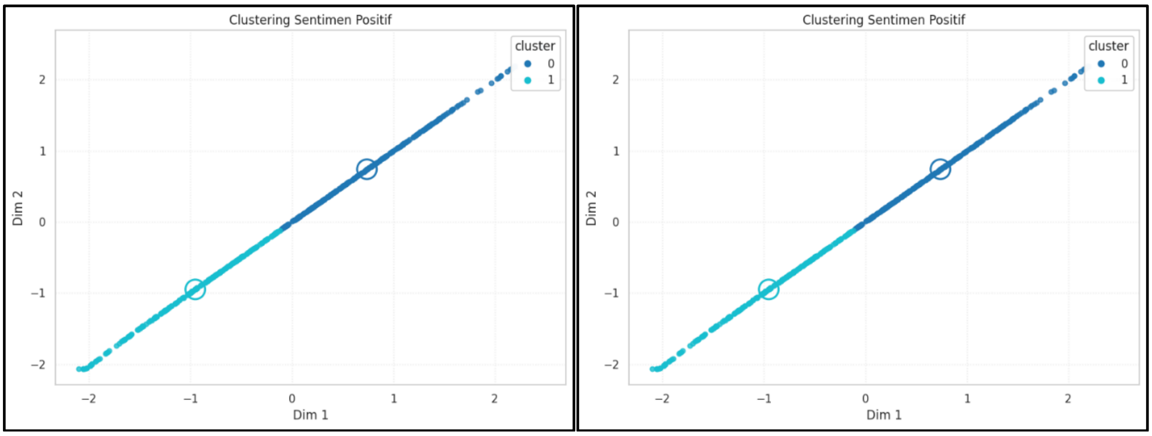

Figure 5. Data distribution in clusters with negative and positive sentiment

The final stage involves extracting pros and cons using the TF-IDF method. Within the {polarity × cluster} combination, dominant terms in positive reviews are treated as strengths (pros), while dominant terms in negative reviews are treated as weaknesses (cons). The mapping of these results is presented in Figures 6 and 7, which clearly highlight the keywords that emerge in each cluster.

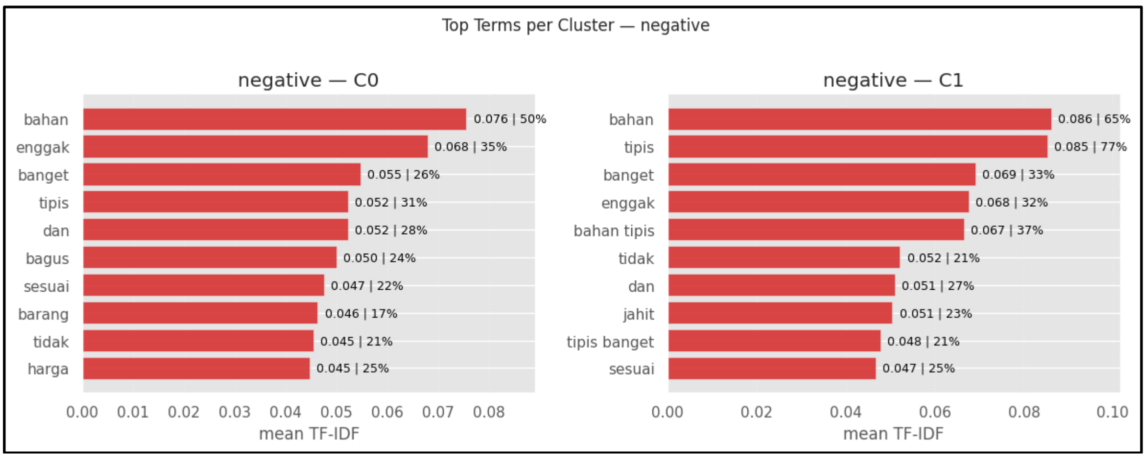

Figure 6. The words that appear in the two negative clusters.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

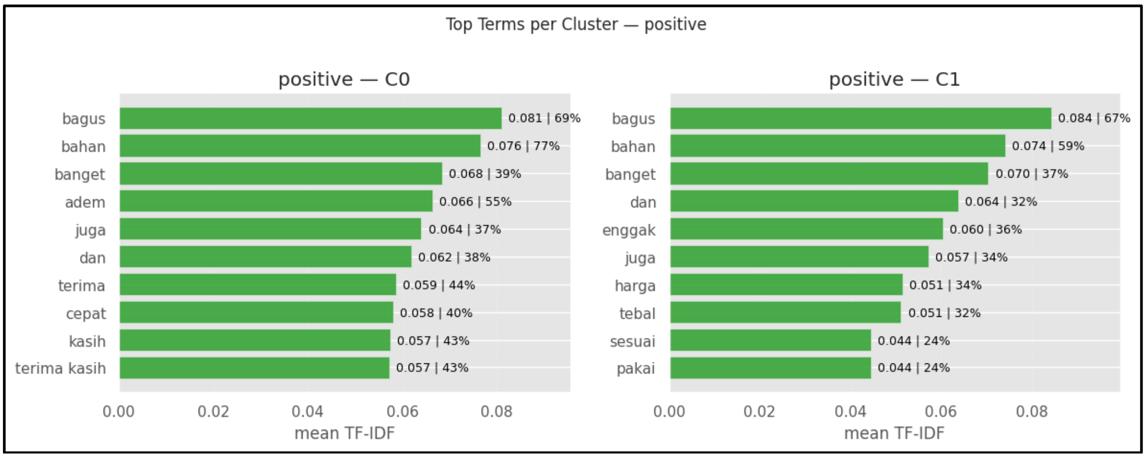

Figure 7. The words that appear in the two positive clusters.

Thus, the pipeline of Word2Vec → CNN-1D → Deep K-Means → TF-IDF has proven effective for automatically extracting product strengths and weaknesses from customer reviews, supported by quantitative metrics, visual evaluation, and the analysis of dominant terms.To address the research question, 'To what extent are deep learning methods effective in clustering and analyzing linguistically diverse customer reviews on Shopee to obtain relevant and socially meaningful product insights?', a series of research activities was conducted. The detailed explanation is as follows.

The effectiveness of deep learning methods in clustering and analyzing customer reviews on Shopee is evident from the classification and clustering results. At the classification stage, the Word2Vec-based CNN-1D demonstrated high performance, achieving an accuracy and F1-macro score of 0.94. Table 1 shows that precision and recall for both positive and negative classes were well-balanced, further supported by visual evidence from the confusion matrix (Figure 5) and the ROC and Precision–Recall curves (Figure 6). These results confirm that polarity labeling can be reliably used as the foundation for further analysis.

Table 1. Classification performance of each class

Class	Precision	Recall	F1
Negative	0.93	0.96	0.94
Positive	0.96	0.94	0.94

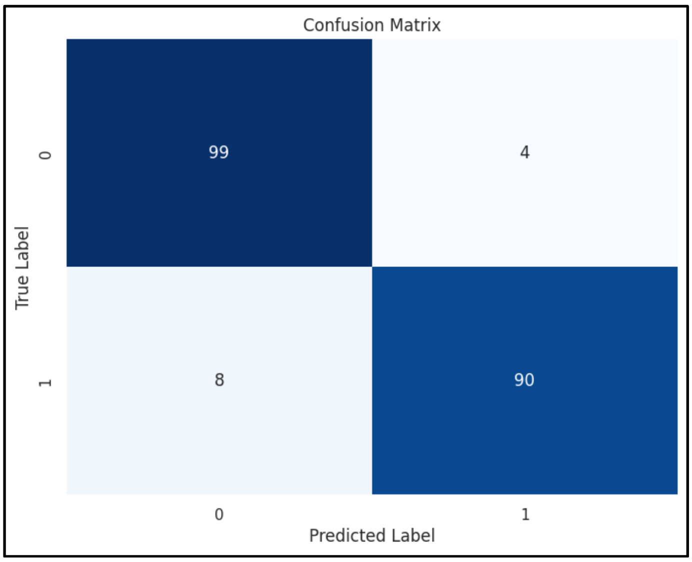

Figure 8. The words that appear in the two positive clusters

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

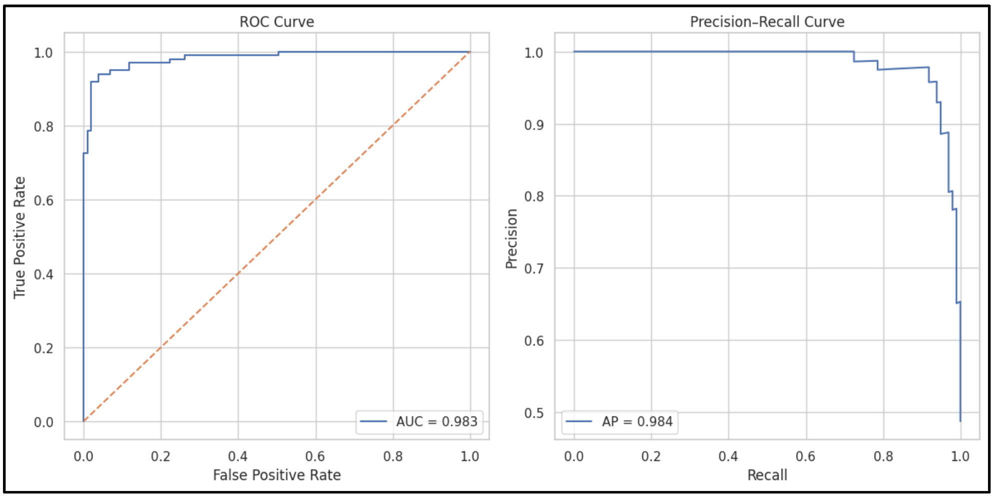

Figure 9. ROC Curve and Precision-Recall

Meanwhile, at the aspect clustering stage, the Deep K-Means model consistently formed compact and well-separated clusters. Evaluation using the Silhouette Score and DBI across different cluster numbers supported the choice of k = 2 for both positive and negative corpora. The best Silhouette and DBI values are presented in Table 2 below, while the cluster data distribution is visualized in Figure 2 above. The semantic consistency of the clusters was further reinforced through an inspection of dominant words based on TF-IDF, shown in Figures 7 and 8. This indicates that the pipeline not only accurately determines polarity but also effectively groups reviews into relevant aspects.

Table 2. Silhouette Score Value and Davies–Bouldin Index for each class

Klaster	Kelas Negatif		Kelas Positif	
	Silhouette	DBI	Silhouette	DBI
1	0.576	0.577	0.597	0.535
2	0.573	0.508	0.554	0.540
3	0.553	0.513	0.540	0.534
4	0.537	0.533	0.551	0.528
5	0.544	0.506	0.526	0.539
6	0.550	0.503	0.532	0.546
7	0.543	0.511	0.553	0.492

These findings indicate that the deep learning methods employed in this study are effective in classifying and clustering customer reviews. The model not only excels in terms of metrics, achieving high accuracy and producing well-structured clusters, but also provides informative summaries of pros and cons. The pipeline of Word2Vec, CNN-1D, and Deep K-Means, supported by TF-IDF, can therefore be regarded as a reliable and practically valuable approach for extracting relevant product insights from customer reviews in the Shopee marketplace.

4. DISCUSSIONS

The results of this study demonstrate that deep learning methods are effective in both classifying sentiment and extracting aspect-level insights from linguistically diverse customer reviews on Shopee. The Word2Vec-based 1D CNN achieved consistently high accuracy (0.93–0.94) and balanced F1-scores across positive and negative classes, confirming the model's robustness in handling polarity detection. The ROC and Precision–Recall curves further verified that performance was stable under varying decision thresholds, suggesting resilience to class imbalance and variability in linguistic expression.

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

These outcomes highlight the ability of CNN models to capture local linguistic features crucial for sentiment analysis in e-commerce contexts. These findings are in line with previous studies that have shown CNN-based architectures to be highly effective for sentence-level sentiment classification. Kim (2014) demonstrated that CNNs outperform traditional models, such as SVM and logistic regression, in capturing the semantic features of short texts, which supports the results obtained in this study. Similarly, Zhang et al. (2018) highlighted the effectiveness of Word2Vec embeddings in improving classification accuracy by representing semantic proximity between words, particularly in informal digital communication.

The clustering stage using Deep K-Means further supports the utility of deep learning in aspect-based sentiment analysis. The moderate Silhouette and Davies–Bouldin Index values (Silhouette ~0.6, DBI ~0.5) indicate adequate cohesion and separation between clusters, consistent with benchmarks reported in related studies. For instance, Xie et al. (2016) proposed Deep Embedded Clustering (DEC), showing that autoencoder-based clustering achieves more meaningful latent representations compared to traditional K-Means alone. More recently, Xu and Tian (2020) applied deep clustering to customer review data and reported similar levels of performance in distinguishing product aspects, aligning with the findings of this study.

From a sociolinguistic perspective, the results confirm that deep learning models are capable of adapting to the linguistic variation commonly found in Indonesian e-commerce reviews. This supports earlier work by Purwarianti and Crisdayanti (2019), who observed that sentiment analysis in Indonesian requires handling informal registers, abbreviations, and code-switching. The balanced classification results in this study suggest that Word2Vec embeddings combined with CNN are sufficiently sensitive to these sociolinguistic features, enabling reliable sentiment detection even in noisy, user-generated text. By integrating Word2Vec, CNN, and Deep K-Means, this study extends prior research and demonstrates how deep learning methods can move beyond polarity classification to uncover structured, actionable insights. The identification of strengths (e.g., durability, design, ease of use) and weaknesses (e.g., material quality, packaging, delivery issues) reflects not only the technical effectiveness of the methods but also their practical value for sellers seeking to improve product quality and customer satisfaction in digital marketplaces.

5. CONCLUSION

Based on the findings, it can be concluded that integrating Word2Vec, convolutional neural networks (CNN), and deep k-means is effective in analyzing Shopee customer reviews, which exhibit significant linguistic variation. The word2vec-cnn model achieved high accuracy with balanced F1-scores across positive and negative classes, demonstrating robustness in detecting polarity despite the presence of slang, abbreviations, and code-switching in the text. The aspect clustering stage, utilizing deep k-means, generated an optimal configuration with two clusters for each sentiment, resulting in silhouette scores and Davies–Bouldin index values that indicate sufficient intra-cluster cohesion and clear inter-cluster separation. The clustering results revealed that positive reviews primarily highlighted aspects such as product design, durability, and ease of use, while negative reviews emphasised material quality, packaging, and delivery time. Thus, this approach not only classifies sentiment effectively but also provides structured insights into the product's strengths and weaknesses. From a sociolinguistic perspective, the findings demonstrate that deep learning models can effectively accommodate the linguistic diversity of Indonesian e-commerce discourse, generating socially meaningful and actionable

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

insights. This study practically contributes to helping sellers improve product quality and marketing strategies by offering a deeper understanding of customer perception.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest between the authors or with the research object in this paper.

ACKNOWLEDGEMENT

The authors would like to send their gratitude to DPPM for fully funding this article.

REFERENCES

[1] S. Bozkurt, D. Gligor, L. D. Hollebeek, and C. Sumlin, "Understanding the effects of firms' unresponsiveness on social media toward customer feedback on customers' engagement: the impact of ethnicity," *Journal of Research in Interactive Marketing*, Jan. 2024, doi: 10.1108/JRIM-09-2023-0317.

[2] A. D. Rachmadanty, A. Audiayani Muhtar, and A. Agustina, "Management Examining the Impact of Gamification and Customer Experience on Customer Loyalty in E-commerce: Mediating Role of Customer Satisfaction," *Journal of Enterprise and Development (JED)*, vol. 7, no. 1, p. 2025, 2025.

[3] A. Yuille and U. Mir, "Causal Inference in Customer Feedback Analysis: A Benchmarking Approach with LLMs," *Journal of Economic in Globalization*, vol. 20, no. 3, pp. 12–24, 2025, doi: 10.13140/RG.2.2.11856.52486.

[4] N. Ruytenbeek, S. Decock, and I. Depraetere, "The impact of linguistic choices and (para-)linguistic markers on the perception of Twitter complaints by other customers: an experimental approach," *Journal of Politeness Research*, vol. 19, no. 1, pp. 87–122, 2023, doi: doi:10.1515/pr-2021-0031.

[5] A. Rajuroy, "Sentiment Analysis of Amazon Product Reviews Using Machine Learning Approaches", [Online]. Available: https://www.researchgate.net/publication/394832633

[6] K. Ollivier, C. Boldrini, A. Passarella, and M. Conti, " Unveiling Cognitive Constraints in Language Production: Extracting and Validating the Active Ego Network of Words ," *IEEE Trans Comput Soc Syst*, pp. 1–13, Jul. 2024, doi: 10.1109/tcss.2024.3419565.

[7] H. K. Smith, "Social Media Marketing and Digital Presence: Entrepreneurial Approaches in Oman's Logistics Sector," *Journal of Business Digital*, vol. 12, no. 1, pp. 146–148, 2025, [Online]. Available: https://www.researchgate.net/publication/389451377

[8] S. Salloum, O. Khasoneh, R. Abousamra, and A. AlHamad, *Generative AI in Advertising and Marketing: A BERT-Based Analysis of Text Descriptions for Product Advertisement*. Willey, 2025. doi: 10.1007/978-3-031-89175-5_17.

[9] Fahry and E. Utami, "Optimizing Sentiment Analysis Of Product Reviews On Marketplace Using A Combination Of Preprocessing Techniques, Word2vec, And Convolutional Neural Network Optimisasi Analisis Sentimen Ulasan Produk Pada Marketplace Dengan Kombinasi Teknik Preprocessing, Word2vec, Dan Convolutional Neural Network," *Jurnal Teknik Informatika (JUTIF)*, vol. 4, no. 1, pp. 101–107, 2023, doi: 10.20884/1.jutif.2023.4.1.815.

[10] R. Bharathi, R. Bhavani, and R. Priya, "Leveraging deep learning with sentiment analysis for Online Book reviews polarity classification model," *Multimed Tools Appl*, vol. 84, pp. 29843–29862, Sep. 2024, doi: 10.1007/s11042-024-20369-7.

[11] J. A. Josen Limbong, I. Sembiring, K. Dwi Hartomo, U. Kristen Satya Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Neighbor Analysis of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naive Bayes and K-Nearest Neighbor Method," *Journal of Multi*, 2024, doi: 10.25126/jtiik.202294960.

[12] M. Schoop, "Negotiation communication revisited," *Cent Eur J Oper Res*, vol. 29, no. 1, pp. 163–176, Mar. 2021, doi: 10.1007/s10100-020-00730-5.

[13] E. Diegoli, “A corpus-assisted analysis of indexical signs for (im)politeness in Japanese apology-like behaviour,” *Journal of Politeness Research*, vol. 20, no. 2, pp. 427–453, 2024, doi: doi:10.1515/pr-2022-0002.
[14] I. Hübscher, C. Sánchez-Conde, J. Borràs-Comes, L. Vincze, and P. Prieto, “Multimodal mitigation: how facial and body cues index politeness in Catalan requests,” *Journal of Politeness Research*, vol. 19, no. 1, pp. 1–29, 2023, doi: doi:10.1515/pr-2020-0033.
[15] Z. Yang, Y. Du, D. Liu, K. Zhao, and M. Cong, “A human-robot interaction system for automated chemical experiments based on vision and natural language processing semantics,” *Eng Appl Artif Intell*, vol. 146, p. 110226, 2025, doi: https://doi.org/10.1016/j.engappai.2025.110226.
[16] M. B. M. Hansen, “Social meaning as Hearer’s Meaning: Integrating social meaning into a general theory of meaning in communication,” *J Pragmat*, vol. 241, pp. 81–91, May 2025, doi: 10.1016/j.pragma.2025.03.010.
[17] F. González, M. Torres-Ruiz, G. Rivera-Torruco, L. Chonona-Hernández, dan R. Quintero, “A Natural-Language-Processing-Based Method for the Clustering and Analysis of Movie Reviews and Classification by Genre,” *Mathematics*, vol. 11, no. 23, p. 4735, 2023, doi:10.3390/math11234735.
[18] Q. Xu, “Text clustering based on pre-trained models and deep embedding clustering,” *Frontiers in Computational Neuroscience*, 2024, doi: 10.3389/fncom.2023.1334436.
[19] P. Vijayaragavan et al., “Transforming sentiment analysis for e-commerce product reviews: Hybrid deep learning model with an innovative term weighting and feature selection,” *Information Fusion*, 2024
[20] J. A. Aguilar-Moreno, P. R. Palos-Sanchez, dan R. Pozo-Barajas, “Sentiment analysis to support business decision-making. A bibliometric study,” *AIMS Mathematics*, vol. 9, no. 2, pp. 4337-4375, 2024, doi: 10.3934/math.202421
[21] S. Al-Hadhrami, T. Vinko, T. Al-Hadhrami, F. Saeed, dan S. N. Qasem, “Deep learning-based method for sentiment analysis for patients’ drug reviews,” *PLoS ONE*
[22] M. J. Verdú et al., “Clustering of LMS Use Strategies with Autoencoders,” *Applied Sciences*, 2023.
[23] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” arXiv preprint arXiv:1408.5882, 2014. [Online]. Available: https://arxiv.org/abs/1408.5882
[24] A. H. M. Azam, “Asymmetric Effect of Market Sentiment on Banking: A Nonlinear ARDL Approach,” *Jurnal Ekonomi Malaysia*, vol. 58, no. 1, pp. 1– 15, 2024, doi: 10.17576/JEM-2024-5801-08.
[25] S. Bozkurt, D. Gligor, L. D. Hollebeek, and C. Sumlin, “Understanding the effects of firms’ unresponsiveness on social media toward customer feedback on customers’ engagement: the impact of ethnicity,” *Journal of Research in Interactive Marketing*, Jan. 2024, doi: 10.1108/JRIM-09-2023- 0317
[26] A. Gülbaşı and E. Taşkın, “The Two Faces Of E-Commerce: A Comparison Of ECommerce Platforms And Social Commerce,” *Dumlupınar Üniversitesi İİBF Dergisi*, vol. 2, no. 14, pp. 71–82, Dec. 2024, doi: 10.58627/dpuiibf.1535413.
[27] N. L. Putri, B. Warsito, and B. Surarso, “Pengaruh Klasifikasi Sentimen Pada Ulasan Produk Amazon Berbasis Rekayasa Fitur dan K-Nearest Negihbor,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 65–74, Feb. 2024, doi: 10.25126/jtiik.20241117376
[28] D. Safryda Putri et al., “Analisis Sentimen dan Pemodelan Ulasan Aplikasi AdaKami Menggunakan Algoritma SVM dan KNN,” *Jurnal Swabumi*, vol. 1, no. 3, 2025.
[29] M. Irfani and S. Khomsah, “Analisis Sentimen Berbasis Aspek pada EDOM Pembelajaran Menggunakan Metode CNN dan Word2vec,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 12, no. 3, p. 413, Jul. 2024, doi: 10.26418/justin.v12i3.75610.
[30] E. Y. Hidayat and D. Handayani, “Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 3, pp. 153–163, Jan. 2023, doi: 10.25077/teknosi.v8i3.2022.153-163.
[31] A. Hanafiah et al., “Sentimen Analisis Terhadap Customer Review Produk Shopee Berbasis Wordcloud Dengan Algoritma Naïve Bayes Classifier Sentiment Analysis Of Customer Reviews Of Shopee Products Based On Wordcloud Using Naïve Bayes Classifier

Vol. 6, No. 6, December, 2025, Page. 5489-5502
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.6.5340

Algorithm," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, pp. 23–31, 2023.

[32] G. Andrasthea and H. Februariyanti, "Analisis Sentimen Ulasan Herborist Sistem Pengambilan Keputusan Menggunakan Klasifikasi Neighbor dan TF- ID," *Jurnal Swabumi*, vol. 12, no. 2, pp. 176–181, 2024.

[33] M. H. Mekkawi, "The Implication of AI on E-Commerce," *International Journal of Legal and Social Order*, vol. 4, no. 1, Nov. 2024, doi: 10.55516/ijlso.v4i1.201.

[34] J. A. Putra, A. Dharmawan, and J. Gondohanindijo, "Sentimen Analysis Digitalent Mobile Application Using Naïve Bayes And Svm With Tf-Idf Fitur Extraction," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 7, no. 4, 2024.

[35] A. Kumar, D. R. P. M. Vincent, K. Srinivasan, and C. Y. Chang, "Deep Convolutional Neural Network based Feature Extraction with optimized Machine Learning Classifier in Infant Cry Classification," in 2020 International Conference on Decision Aid Sciences and Application, DASA 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 27–32. doi: 10.1109/DASA51403.2020.9317240.

[36] S. Purkovic et al., "Audio analysis with convolutional neural networks and boosting algorithms tuned by metaheuristics for respiratory condition classification," Journal of King Saud University -Computer and Information Sciences, vol. 36, no. 10, Dec. 2024, doi: 10.1016/j.jksuci.2024.102261

[37] L. F. A. O. Pellicer, T. M. Ferreira, and A. H. R. Costa, "Data augmentation techniques in natural language processing," Appl Soft Comput, vol. 132, Jan. 2023, doi: 10.1016/j.asoc.2022.109803.

[38] A. A. Abdulnassar and L. R. Nair, "Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model," Meas. Sensors, vol. 25, no. 100666, 2023, doi: 10.1016/j.measen.2023.10066

[39] A. Alamsyah et al., "Customer Segmentation Using the Integration of the Recency Frequency Monetary Model and the K-Means Cluster Algorithm," Sci. J. Informatics, vol. 9, no. 2, pp. 189–196, 2022, doi: 10.15294/sji.v9i2.39437.

[40] M. Al Ghifari and W. T. Harsanti Putri, "Clustering Courses Based On Student Grades Using K-Means Algorithm With Elbow Method For Centroid Determination," Inf. J. Ilm. Bid. Teknol. Inf. dan Komun., vol. 8, no. 1, pp. 42–46, 2023, doi: 10.25139/inform.v8i1.4519.