Vol. 6 No. 3, 2025, pp. 196-206 **Jurnal Pepadun**
© 2025 Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)

Sentiment Analysis of Gamification in E-Commerce Applications Using a Hybrid CNN-LDA

[*1]Rayyandra Rezky, [2]Ulfa Khaira, and [3]Mutia Fadhila Putri

[1,2,3]Faculty of Science and Technology, Jambi University
Jl. Jambi - Muara Bulian KM. 15, Mendalo Darat, Muaro Jambi Regency, Jambi, Indonesia
e-mail: **[*1]rezkyrayy@gmail.com**, [2] ulfakhaira@unja.ac.id, [3]mutia.fadhila@unja.ac.id

Abstract *- This study investigates public sentiment toward gamification features in e-commerce platforms, focusing on user opinions expressed on social media platform X. As gamification strategies like Shopee Candy and Tokopedia Quiz become more common, understanding user perceptions is crucial for improving user experience. This research adopts a hybrid approach using Convolutional Neural Network (CNN) for sentiment classification and Latent Dirichlet Allocation (LDA) for topic modeling. A total of 765 user comments were collected using Tweet Harvest and processed through standard text preprocessing techniques. The CNN model achieved strong performance, with 81% accuracy, 86% precision, 89% recall, and an F1-score of 88%. LDA analysis revealed key terms in each sentiment group—positive sentiments centered on words like "menang" and "hibur," while negative sentiments included "scam" and "capithaha." The results demonstrate the effectiveness of combining CNN and LDA for analyzing sentiment and extracting dominant themes in informal online discussions. These insights can guide e-commerce platforms in refining gamification strategies that align with user expectations, helping to enhance engagement while addressing user concerns.*

Keywords**:** *Convolutional Neural Network; E-commerce; Gamification; Latent Dirichlet Allocation; Sentiment Analysis.*

1. INTRODUCTION

The rapid advancement of digital technology has fundamentally transformed global business practices, with electronic commerce (e-commerce) emerging as one of the most disruptive innovations of the digital era [1]. As a comprehensive digital ecosystem encompassing transactions, data exchange, and digital payments, e-commerce has not only revolutionized traditional retail models but also dramatically reshaped consumer behavior worldwide. In Indonesia, this transformation has been particularly remarkable, with the sector's gross merchandise value (GMV) reaching USD 62 billion in 2023 and projected to maintain strong growth momentum. This expansion is fueled by increasing internet penetration, widespread smartphone adoption, and changing consumer preferences that favor the convenience, competitive pricing, and accessibility of online shopping platforms [2].

In this dynamic landscape, text mining has emerged as a crucial analytical tool for extracting valuable insights from the vast amounts of unstructured user-generated content on e-commerce platforms and social media. As a specialized branch of data mining, text mining employs natural language processing (NLP) and machine learning techniques to transform raw textual data into meaningful patterns and actionable knowledge [3]. Text mining was initially utilized as an early approach to business intelligence, which was subsequently advanced into the development of a text mining prototype [4]. The application of text mining in e-commerce has become increasingly sophisticated, enabling businesses to analyze product reviews, customer feedback, and social media conversations at scale. This capability is particularly valuable for understanding complex consumer behaviors and preferences that would otherwise remain hidden in unstructured data [5].

With the rise of online shopping driven by convenience and cost savings, e-commerce platforms increasingly rely on recommendation systems and gamification strategies to enhance user engagement and stay competitive [6]. By incorporating game-like elements such as points, badges, leaderboards, and rewards into non-game contexts, platforms like Shopee, Tokopedia, and Lazada aim to enhance user interaction and foster loyalty [7].

Notable implementations include Shopee's "Shopee Candy" rewards system and Tokopedia's interactive "Tokopedia Play" features, which have demonstrated measurable impacts on user retention and transaction frequency. Empirical studies indicate that well-executed gamification strategies can increase transaction frequency by up to 25% and boost platform visits by 40% [8].

However, user responses to these gamification features show significant variation. While some consumers enthusiastically engage with these game-like elements, others express frustration stemming from perceived manipulative design, repetitive mechanics, or unclear reward structures. This dichotomy creates a critical need for sophisticated sentiment analysis techniques that can accurately interpret and categorize user opinions from the vast amounts of unstructured text data generated across social media platforms and review sites [9][10].

One notable example is the application of sentiment analysis to detect cyberbullying activities within the comment sections of social media platform X. Using a dataset obtained from Kaggle, the research achieved an accuracy rate of approximately 80% [11]. Beyond this, sentiment analysis has evolved into more advanced approaches, such as the integration of CNN and LSTM methods to analyze public comments on the Sirekap system, which has drawn significant attention from Indonesian society [12]. In its earlier stages, however, sentiment analysis primarily relied on more traditional approaches.

Traditional approaches to sentiment analysis, including lexicon-based methods and Naïve Bayes classifiers, often struggle with the challenges posed by informal language use, slang, and contextual nuances prevalent in user-generated content [13]. These limitations become particularly apparent when analyzing colloquial expressions and sarcastic remarks common in social media discussions about gamification features. For instance, phrases like "*game Shopee bikin nagih tapi hadiahnya receh*" ("Shopee's game is addictive but the rewards are meager") present significant interpretation challenges for conventional sentiment analysis tools [14].

To address these limitations, recent advancements in text mining and machine learning have introduced more sophisticated analytical approaches. Convolutional Neural Networks (CNNs) have shown particular promise in sentiment analysis applications due to their ability to capture hierarchical patterns and semantic relationships in textual data. When combined with Latent Dirichlet Allocation (LDA) for topic modeling, this hybrid approach enables researchers to not only determine sentiment polarity but also identify the specific aspects of gamification that drive user satisfaction or frustration [15].

While Convolutional Neural Networks (CNNs) excel at classifying sentiment polarity, their standalone application fails to identify the thematic drivers of these sentiments. The incorporation of Latent Dirichlet Allocation (LDA) is therefore essential, as it provides the critical contextual layer by discovering and classifying the underlying topics within the text. Without this integration, the analysis lacks actionable insight; one might ascertain *that* users are dissatisfied but remain unable to identify what specific aspects are causing frustration, thereby severely limiting the practical applicability of the findings [16].

The primary objective of this research is to analyze user sentiment toward gamification within non-gaming applications, specifically E-Commerce platforms. Beyond a quantitative model, this study aims to provide actionable insights for the future development of gamification features. It will evaluate potential strategies and their anticipated outcomes, with the goal of enhancing user engagement and ultimately driving revenue growth for E-Commerce businesses.

2. RESEARCH METHODOLOGY

There are few parts from text mining that being used on this research like data mining that focused on knowledge in discovery [17], Text Extraction That focused on extracting the text / data that are indicate in a picture or a document [18]. The primary methods used were Convolutional Neural Network (CNN) for sentiment classification and Latent Dirichlet Allocation (LDA) for topic modeling. The entire research process was conducted from September 2024 to June 2025 at the Information Systems Department, Universitas Jambi.

© 2025 The Authors | doi: 10.23960/pepadun.v6i3.281

2.1 Data Collection

The dataset comprised 765 user comments scraped from social media platform X (formerly Twitter) using the Tweet Harvest tool. Keywords such as #shopeecocoki, #tokopediaplay, and other gamification-related tags were used to filter relevant posts [19]. The raw dataset preserved the complete textual content along with 15 additional metadata fields, providing rich contextual information for subsequent analysis.

Table 1. Collected data.

No	Conversation_Id_Str	Created_At	Full_Text	Id_Str	Reply_Count	Username
1.	10847	Tue Oct 22 2024	Ah ini shopee capit katrok banget deh #gambarjelek #shopeecapit	1084	9	Saipul_ganteng12
2.	20125	Mon jan 26 2025	Dapet 2 voucher 50% dari cocoki nich asikk #terimakasihshopee #shopeecocoki	2004	23	JeniBlekping

The dataset also holds a lot of information such as Conversation_id_str, Created_at, Favorite_count, Full_text, Id_str, Image_url, In_reply_to_screen_name, Lang, Location, Quote_count, Reply_count, Retweet_count, Tweet_url, User_id_str and Username, with Table 1 only presenting a few examples from the dataset.

2.2 Dropping Column

After the data has been scraped, next step is preprocessing the data. Preprocessing data is a process of removing and adjusting data so it would do as the research need, it can be also called as cleaning data which has almost the same purpose. Removing the unnecessary information [20] such as Conversation_id_str, Created_at, Favorite_count, Id_str, Image_url, In_reply_to_screen_name, Lang, Location, Quote_count, Reply_count, Retweet_count, Tweet_url, User_id_str and Username. The output after this process is presented in Table 2.

Table 2. After dropping column.

Full Text
Siapasih yang ciptain cocoki !! Bikin gila aja untung gak masuk RSJ
waahh padahal hari ni nyerah bgtt mana td sempat ke refresh sampe blg anjir pdhl mau menang (pdhl masii byk hiks) pas balik ke apk ternyata ngelanjutin ga ngulang dari awal eh beneran menang dong ketemu ni passion
masih penasaran mecahin game misteri cocoki shopee satuu iniii
Aku ttp akan memperjuangkan game ini disaat arhan tetiba kawin sama ziza
pejuang cocoki shopee game nya emng bikin candu smpe rela tengah malem masih aja main :v

2.3 Data Labelling

The remaining data is then prepared for the labeling process, which is a critical stage in supervised learning-based sentiment analysis. In this research, a manual labeling technique is employed to ensure high accuracy and contextual understanding. Instead of relying solely on automated or lexicon-based sentiment scoring—which may fail to capture nuances in informal language—the annotation is performed manually with assistance from language expertise from Jambi Provincial Language Center, who is familiar with the cultural, emotional, and linguistic subtleties of Indonesian social media content [21].

Each tweet is carefully examined and classified into one of two sentiment categories: positive or negative. To minimize subjectivity, several measures were implemented. **First**, expert involvement was crucial in resolving ambiguities such as sarcasm, mixed sentiment phrases, and the use of slang or local expressions that are often missed by automated systems. **Second**, the annotation team followed a standardized set of annotation

© 2025 The Authors | doi: 10.23960/pepadun.v6i3.281

guidelines, which defined clear criteria and examples for labeling, ensuring consistency across annotators. **Third**, the labeling process incorporated cross-checking, where multiple annotators reviewed a subset of the data and discussed discrepancies until consensus was reached. Resulting final data has been labeled based on the sentiment, these steps helped reduce personal bias, improve reliability, and ensure that sentiment labels accurately reflected the intended meaning of the content as shown in Table 3.

Table 3. After manual labelling.

Full Text	Label
Siapasih yang ciptain cocoki !! Bikin gila aja untung gak masuk RSJ	0
waahh padahal hari ni nyerah bgtt mana td sempat ke refresh sampe blg anjir pdhl mau menang (pdhl masii byk hiks) pas balik ke apk ternyata ngelanjutin ga ngulang dari awal eh beneran menang dong ketemu ni passion	0
masih penasaran mecahin game misteri cocoki shopee satuu iniii	1
Aku ttp akan memperjuangkan game ini disaat arhan tetiba kawin sama ziza	1
pejuang cocoki shopee game nya emng bikin candu smpe rela tengah malem masih aja main :v	1

2.4 Processing Data

After the data was labeled, it proceeded through several Natural Language Processing (NLP) stages. The first step was case folding, which involved converting all uppercase letters into lowercase to maintain consistency across textual data [22]. Following this, a Num2Word conversion was applied, transforming numeric values into their corresponding textual representations (e.g., "1" became "one"). Words containing embedded digits (such as “hello12”), single-character tokens or those with fewer than two letters (e.g., “co”), and excessively long words (e.g., nonsensical strings like “kajsasoneasjdamnnn”) were removed to reduce noise. Punctuation marks were also eliminated as part of this cleaning process.

Although these actions are often grouped under stop word filtering, they primarily aim to refine the dataset by removing linguistically irrelevant or malformed tokens. After these cleansing steps, tokenization was carried out, in which each sentence was split into individual word units or tokens, preparing the data for further modeling and analysis.

2.5 Convolutional Neural Network and Parameters

The model was constructed using a sequential architecture designed to progressively transform raw text into a sentiment classification. The first layer was an Embedding Layer. This layer is crucial for converting each tokenized word into a dense numerical vector of a fixed dimension (e.g., 100, 200). The parameters for this layer were set with input_dim equal to the vocabulary size, output_dim defining the embedding vector size, and input_length set to the maximum sequence length of the tokenized texts [23]. This transformation allows the model to capture semantic relationships between words. The output from the embedding layer was passed to a One-Dimensional Convolutional Layer (Conv1D). This layer acts as a feature detector by applying multiple filters to sliding windows of the embedded word sequence. Key parameters for this layer include the number of filters (e.g., 64 or 128), which determines the depth of the output, and the kernel_size (e.g., 3 or 5), which defines the number of words each filter considers concurrently [24]. The ReLU (Rectified Linear Unit) activation function was used to introduce non-linearity.

Subsequently, a GlobalMaxPooling1D Layer was applied to the convolutional output. This layer reduces the dimensionality of the feature maps by extracting the maximum value from each filter's output sequence. This operation highlights the most prominent feature detected by each filter, effectively capturing the strongest emotional signals in the text for classification while improving computational efficiency. The condensed feature vector from the pooling layer was then fed into a Dense Layer with a ReLU activation function [25]. This fully connected layer learns non-linear combinations of the high-level features extracted by the previous layers. Finally, the output was directed to a Dense Output Layer with a single unit and a sigmoid activation

© 2025 The Authors | doi: 10.23960/pepadun.v6i3.281

function. This configuration outputs a probability between 0 and 1, representing the confidence that the input text expresses a positive sentiment. The model was compiled with the Adam optimizer and binary cross-entropy loss function for training.

2.6 Topic Modeling Using LDA

Beyond just classifying sentiment, this study also delved into the underlying discussions within the tweets using Latent Dirichlet Allocation (LDA), an unsupervised learning model. LDA operates on the principle that each tweet is a mix of various topics, and each topic is a mix of specific words. To implement this, the preprocessed text was first transformed into a document-term matrix using Count Vectorizer, quantifying word frequencies. LDA was then applied to pinpoint clusters of co-occurring words, effectively uncovering the dominant themes. A separate analysis was conducted for positive and negative sentiment classes to highlight differences in their topic composition.

This topic modeling process yielded distinct word groupings. Positive sentiments were linked to topics featuring words like "*joki*," "*menang*," and "*hibur*," signaling user excitement and engagement, and achieving a strong coherence score of 0.7. In contrast, negative sentiments revealed keywords such as "scam," "*pusing*," and "*gagal*," indicating user frustration or dissatisfaction, with a coherence score of 0.5. Through this detailed topic modeling, the study provided valuable qualitative insights into the specific issues and attractions users experienced with gamified e-commerce features, enriching the overall sentiment analysis.

2.7 Evaluation

The CNN model evaluation in sentiment classification is using standard metrics which is confusion matrix that holds variable such as True Positive, False Positive, True Negative and False Negative. From those variables and going through each math formula, resulting Accuracy, Precision, Recall and F1-Score as presented in Formula 1 to 4.

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

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

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

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

These values were calculated using a confusion matrix generated from the test data, which represented 20% of the total dataset. The use of TensorFlow and scikit-learn libraries facilitated the training and evaluation process efficiently. Meanwhile, the quality of the topic modeling results was validated using coherence scores, a common metric in topic modeling that assesses the degree of semantic similarity between the top words in each topic. A coherence scores above 0.5 was considered acceptable, indicating that the identified topics were interpretable and consistent with user discussions.

3. RESULTS AND DISCUSSION

This research presents results from a sentiment analysis study on gamification features in e-commerce applications using a combination of Convolutional Neural Network (CNN) and Latent Dirichlet Allocation (LDA). The CNN model was evaluated using standard performance metrics, while LDA was used to extract discussion topics based on sentiment groups.

3.1. Convolutional Neural Network Results

The CNN model was trained on a manually labeled dataset of 765 Indonesian tweets, with labels indicating positive and negative sentiment. The evaluation used a train-test split of 80:20. The results are summarized in Table 4, which shows that the CNN model achieved an accuracy of 81%, precision of 86%, recall of 89%, and F1-score of 88%. These results suggest the model performed reliably on informal and noisy Indonesian-language social media data.

Table 4. CNN overall results.

Parameter	Value
Accuracy	81%
Precision	86%
Recal	89%
F1-Score	88%

The conclusion drawn by the module is based on a series of validation tests conducted to ensure the accuracy and reliability of the model's output. These tests include the use of evaluation metrics such as accuracy, precision, recall, and F1-score on a reserved test dataset. By applying cross-validation techniques [26], the model's consistency and generalizability were further confirmed as shown in Table 5, whereas Table 4 highlights key performance indicators used in sentiment analysis. These tables serves as reference point to justify the model's final predictions and demonstrates the validity of the CNN-based classification approach on Indonesian-language social media data.

Table 5. CNN test results.

No	Accuracy	Precision	Recall	F1-Score
1	0.8170	0.8560	0.9145	0.8843
2	0.8105	0.8607	0.8974	0.8787
3	0.7974	0.8707	0.8632	0.8670
4	0.8039	0.8480	0.9060	0.8760
5	0.8235	0.8516	0.9316	0.8898
6	0.8170	0.8504	0.9231	0.8852
7	0.8301	0.8583	0.9316	0.8934
8	0.8170	0.8504	0.9231	0.8852
9	0.8170	0.8504	0.9231	0.8852

3.2. Latent Dirichlet Allocation Results

Following sentiment classification, LDA was applied to uncover key discussion themes within each sentiment group. Texts labeled as positive or negative were processed separately to extract meaningful word groupings. In the positive sentiment group, LDA generated topics featuring keywords such as "*joki*," "*menang*," "*hibur*" and "main." These topics suggest that users were highly engaged by the gamified features and found enjoyment or satisfaction in gameplay and rewards. The coherence score for this topic model was 0.7, indicating well-structured and interpretable topics. In contrast, the negative sentiment group revealed topics with keywords such as "*scam*," "*capithaha*," "*gagal*," and "*ngeselin*." These indicate user frustration with various aspects of the gamified features, including possible bugs, unfair reward mechanisms, or unclear instructions. The coherence score also giving a valid values that saying the result is above 0.5.

Table 6. LDA topics keywords results.

Topics	Keywords	Sentiment	Coherence
1	*joki, menang, hibur, main*	Positive	0.7
2	*scam, capithaha, gagal, ngeselin*	Negative	0.6

The LDA module produced topic results that clearly dominant term for each sentimen. Positive and negative, these results are supported by Table 7 and confirming the coherence score to extract meaningful themes from informal indonesian-language data and enhancing the overal interpretability of the sentiment analysis

Table 7. LDA main topics result.

No	Sentiment	Main Topics (Dominant Word)	Coherence
1	Positive	fruity, bayar, mabar	0.7422
2	Negative	dusta, rugi, tagih	0.7240
3	Positive	bgssst, mantap, candu	0.7910
4	Negative	kesel, sisa, mabar	0.6982
5	Positive	kemarin, kocak, menang	0.5776
6	Negative	sebel, tag, kalah	0.7434
7	Positive	versi, tokennya, fruity	0.7128
8	Negative	scam, ceki, blast	0.7701
9	Positive	blast, fyp, candy	0.6114
10	Negative	fruity, nyawa, tips	0.6518
11	Positive	sbyfess, trik, youngboss	0.7579
12	Negative	malem, blast, pinjem	0.7314
13	Positive	hoki, nyawa, solusi	0.7434
14	Negative	cek, boneka, gabisa	0.7685
15	Positive	youngboss, yay, bubble	0.7865
16	Negative	bantuin, durian, crush	0.5370
17	Positive	durian, candy, kasih	0.7371
18	Negative	rugi, capithaha, crush	0.7795

3.3. Visualization Results

To complement the evaluation of the sentiment classification model, several visualizations were created to provide a clearer, more intuitive understanding of the system's performance and topic distributions. Visualization plays a critical role in making complex data patterns more accessible, especially in interpreting results from deep learning and topic modeling algorithms [27]. This research is using basic heatmap and bar chart that represent the confusion matrix heatmap and the topic distribution charts.

3.4. Confusion Matrix Heatmap Results

A confusion matrix heatmap [28] was generated to visually represent the model's prediction accuracy across both sentiment classes (positive and negative). This heatmap shows the number of true positives, true negatives, false positives, and false negatives in a color-coded grid, enabling quick identification of where the model performs best and where it tends to misclassify. The strong diagonal pattern in the matrix confirms the model's effectiveness in correctly predicting most sentiments.

Model performance, on Figure 1 is evaluated using four key metrics: Accuracy, precision, recall, and F1-score. Accuracy provides an overall measure of prediction correctness, while precision and recall assess class-specific performance, which is especially important for addressing class imbalance [29]. The F1-score balances these two measures, offers more comprehensive evaluation from the model effectiveness.

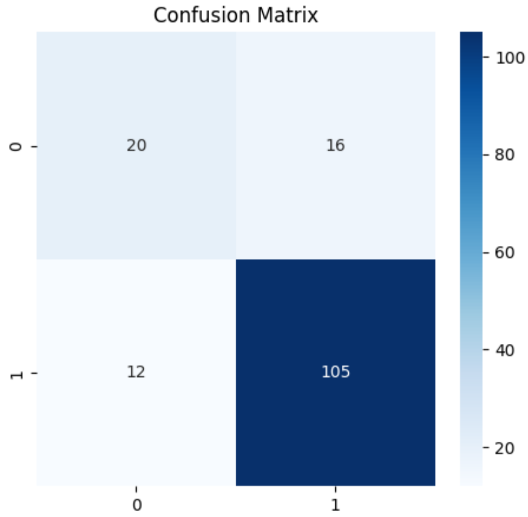

Figure 1. Confusion matrix heatmap.

Accuracy quantifies the proportion of correctly classified samples, both positive and negative, relative to the total number of instances in the test set. It is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{105 + 20}{105 + 20 + 16 + 12} = \frac{125}{153} = 81.7\%$$

For the given confusion matrix, the model achieved an accuracy of 81%, indicating that the majority of predictions were correctly classified. Precision measures the reliability of positive predictions by evaluating how many predicted positives were actual positives. It is defined as:

$$Precision = \frac{TP}{TP + FP} = \frac{105}{105 + 20} = \frac{105}{125} = 86.8\%$$

The model achieved a precision of 86.8%, meaning that most of the samples classified as positive were correct, minimizing dales positives. Recall evaluates the ability of the model to identify all relevant positive case, it goes by:

$$Recall = \frac{TP}{TP + FN} = \frac{105}{105 + 12} = \frac{105}{127} = 89.7\%$$

With a recall of 89.7%, the model successfully detected the majority of true positive cases, although, some were misclassified as negatives. F1-score harmonizes precision and recall into a single metric, defined as:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.868 \times 0.897}{0.868 + 0.897} = 88.2\%$$

The model obtained an F1-score of 88.2%, demonstrating a balanced performance between precision and recall. Overall, these results indicate that the model performs strongly in detecting positive sentiment, though its ability to recognize negative sentiment is weaker, as reflected in a relatively low true negative count (20) compared to the false positives (16). This suggests a tendency toward predicting positive sentiment, which may be influenced by dataset imbalance.

3.5. Topic Distribution Charts Results

Bar charts were used to illustrate the topic distributions obtained through Latent Dirichlet Allocation (LDA). Separate visualizations were generated for positive and negative sentiment. Based on the Figure 2 the image displaying dominance on positive sentiment that reaches 500+ compared to negative sentiment that only reaches 200+. This imbalance gives demonstration a proportion approximately 2,7:1 with positive being the highest. With that the higher proportion of positive instances enables the model to learn more effectively from positive sentiment patterns, which contributes to the strong recall and precision values observed for the positive class in the confusion matrix. Conversely, the relatively smaller number of negative instances limits the model's ability to generalize for this class, resulting in lower specificity and a high number of false positives.

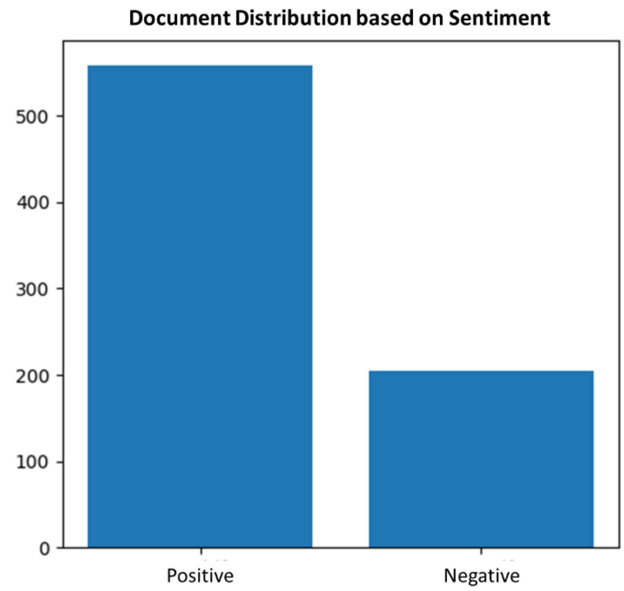

Figure 2. Topic distribution chart.

The presence of class imbalance in sentiment analysis datasets poses a risk of model bias towards the majority class. To mitigate this issue and ensure a more equitable learning process, this project will employ techniques including data-level interventions (oversampling, undersampling, synthetic generation) and algorithm-level adjustments (class weighting). The objective is to improve the generalizability and fairness of the final classification model by balancing the representation of all sentiment classes.

4. CONCLUSIONS

This study demonstrates that integrating Convolutional Neural Networks (CNN) for sentiment analysis with Latent Dirichlet Allocation (LDA) for thematic extraction proves highly effective in evaluating user perceptions of gamification features in e-commerce. The CNN model achieved robust performance metrics, attaining 81% accuracy and an 88% F1-score, confirming its reliability in processing informal Indonesian-language social media content. The LDA analysis effectively uncovered dominant themes within user feedback. Positive responses highlighted elements of fun, achievement, and entertainment, exemplified by terms like "menang," "joki," and "hibur." Conversely, negative feedback centered on issues of complexity, technical glitches, and perceived inequities, with frequent mentions of "scam," "capithaha," and "ngeselin." These insights underscore a critical balance in gamification design: while such features can significantly boost user interaction and satisfaction, their implementation requires meticulous attention to usability and fairness. Suboptimal game mechanics may inadvertently alienate users and damage brand reputation. The methodological approach combining advanced machine learning with topic modeling offers actionable intelligence for product teams, enabling them to refine gamification strategies that truly resonate with their target audience.

However, this study is limited by the relatively small dataset, the focus on a single platform, and the use of binary sentiment categories, which may oversimplify user opinions. Future research should expand to larger and more diverse datasets, apply advanced models such as transformers, and adopt multi-class or aspect-based sentiment analysis to capture more nuanced insights that can further improve the design of gamification strategies in e-commerce.

LITERATURE

[1] V. Jain, B. Malviya, and S. Arya, “An Overview of Electronic Commerce (e-Commerce),” J. Contemp. Issues Bus. Gov., vol. 27, no. 3, 2021, doi: 10.47750/cibg.2021.27.03.090.

[2] D. A. Harahap, “Perilaku Belanja Online Di Indonesia: Studi Kasus,” *JRMSI - J. Ris. Manaj. Sains Indones.*, vol. 9, no. 2, pp. 193-213, 2018, doi: 10.21009/jrmsi.009.2.02.

[3] S. C. Gupta and N. Goyal, “Text Mining: Techniques, Applications and Issues,” vol. 7, no. 11, pp. 283-289, 2024, doi: 10.55524/csistw.2024.12.1.49.

[4] A. Firdaus, W. I. Firdaus, P. Studi, T. Informatika, M. Digital, and P. N. Sriwijaya, “Text Mining,” vol. 13, no. 1, pp. 66-78, 2021. F. Huseynov, “Gamification in E-Commerce,” no. January 2020, pp. 144-161, 2020, doi: 10.4018/978-1-7998-5171-4.ch008.

[5] Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research, International Journal of Information Management, 45, 2019, 191-210.

[6] A. Behl, P. Sheorey, A. Pal, A. Veetil, and S. Singh, "Gamification in E-Commerce: A Comprehensive Review of Literature," J. Electron. Commerce Organ., vol. 18, pp. 1-16, 2020, doi: 10.4018/JECO.2020040101.

[7] M. N. Athoillah and A. Hidayat, “Analysis Of The Role Of Gamification In Influence Shopping Engagement: Shopee E-Commerce Case Study,” vol. 03, no. 04, pp. 212-229.

[8] N. S. Jayawardena, M. Ross, S. Quach, A. Behl, M. Gupta, and L. D. Lang, “Effective Online Engagement Strategies Through Gamification: A Systematic Literature Review and a Future Research Agenda,” J. Glob. Inf. Manag., vol. 30, no. 5, pp. 1-25, 2022, doi: 10.4018/JGIM.290370.

[9] P. Mehta and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” Int. J. Sci. Technol. Res., vol. 9, no. 2, pp. 601-609, 2020.

[10] F. Zamachsari, Gabriel Vangeran Saragih, Susafa’ati, and Windu Gata, “Analysis of Sentiment of Moving a National Capital with Feature Selection Naive Bayes Algorithm and Support Vector Machine,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 4, no. 3, pp. 504-512, 2020, doi: 10.29207/resti.v4i3.1942.

[11] A. Machmud, B. Wibisono, and N. Suryani, “Analisis Sentimen Cyberbullying Pada Komentar X Menggunakan Metode Naïve Bayes,” vol. 5, no. 1, 2025.

[12] H. Handoko, A. Asrofiq, J. Junadhi, and A. S. Negara, “Sentiment Analysis of Sirekap Tweets Using CNN Algorithm,” INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf., vol. 8, no. 2, pp. 312-329, 2024, doi: 10.29407/intensif.v8i2.23046.

[13] H. Hoiriyah, H. Mardiana, M. Walid, and A. K. Darmawan, “Lexicon-Based and Naive Bayes Sentiment Analysis for Recommending the Best Marketplace Selection as a Marketing Strategy for MSMEs,” *J. Pilar Nusa Mandiri*, vol. 19, no. 2, pp. 65-76, 2023, doi: 10.33480/pilar.v19i1.4176.

[14] P. Chaudhari and C. Chandankhede, “Literature survey of sarcasm detection,” *Proc. 2017 Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2017*, vol. 2018-January, no. March 2017, pp. 2041-2046, 2017, doi: 10.1109/WiSPNET.2017.8300120.

[15] F. N. Hikmah, S. Basuki, and Y. Azhar, “Deteksi Topik Tentang Tokoh Publik Politik Menggunakan Latent Dirichlet Allocation (LDA),” *J. Repos.*, vol. 2, no. 4, pp. 415-426, 2020, doi: 10.22219/repositor.v2i4.52.

[16] J. Y. Kim and C. K. Lim, “The Use of Sentiment Analysis and Latent Dirichlet Allocation Topic-Modeling (LDA) on Web Novel Content Quality Factor,” J. Syst. Manag. Sci., vol. 12, no. 2, pp. 242-257, 2022, doi: 10.33168/JSMS.2022.0211.

[17] E. Budiarto, R. Rino, S. Hariyanto, and D. Susilawati, “Penerapan Data Mining Untuk Rekomendasi Beasiswa Pada SD Maria Mediatrix Menggunakan Algoritma C4.5,” Algor, vol. 3, no. 2, pp. 23-34, Mar. 2022, doi: 10.31253/algor.v3i2.1019.

[18] M. Singh, M. S. Kanroo, H. S. Kawoosa, and P. Goyal, "Towards accessible chart visualizations for the non-visuals: Research, applications and gaps," *Computer Science Review*, vol. 48. Elsevier Ireland Ltd, May 01, 2023, doi: 10.1016/j.cosrev.2023.100555.

[19] R. N. Ilyas, S. Mulyani, S. Wiguna, and M. Ramli, "Penggunaan Crawling Data X Dengan Menggunakan Tweet," no. September 2022, pp. 851-856, 2024.

[20] A. P. Joshi and B. V. Patel, "Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process," *Orient. J. Comput. Sci. Technol.*, vol. 13, no. 0203, pp. 78-81, 2021, doi: 10.13005/ojcst13.0203.03.

[21] M. A. Al-Shabi, "Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining," *Int. J. Comput. Sci. Netw. Secur.*, vol. 20, no. 1, pp. 51-57, 2020, [Online]. Available: http://sentiwordnet.isti.cnr.it/.0

[22] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24-33, 2021, doi: 10.57152/malcom.v1i1.20.

[23] I. M. T. Mas and A. E. Karyawati, "Aspect Based Sentiment Analysis Terhadap Ulasan Produk Skincare di E-Commerce Menggunakan CNN-LSTM," vol. 3, pp. 261-270, 2025.

[24] P. A. Aritonang, M. E. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using CNN (Case Study: Peduli Lindungi Application)," *Ultim. Infosys J. Ilmu Sist. Inf.*, vol. 13, no. 1, pp. 54-61, 2022.

[25] A. Al-Sabaawi, H. M. Ibrahim, Z. M. Arkah, M. Al-Amidie, and L. Alzubaidi, "Amended Convolutional Neural Network with Global Average Pooling for Image Classification," pp. 171-180, 2021, doi: 10.1007/978-3-030-71187-0_16.

[26] [D. F. Zhafira, B. Rahayudi, and I. Indriati, "Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube," *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, pp. 55-63, 2021, doi: 10.25126/justsi.v2i1.24.

[27] S. Urologin, "Sentiment analysis, visualization and classification of summarized news articles: A novel approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 616-625, 2018, doi: 10.14569/ijacsa.2018.090878.

[28] B. Paneru, B. Thapa, and B. Paneru, "Sentiment analysis of movie reviews: A flask application using CNN with RoBERTa embeddings," *Syst. Soft Comput.*, vol. 7, no. September 2024, p. 200192, 2025, doi: 10.1016/j.sasc.2025.200192.

[29] G. P. Permana, D. A. Nugraha, and H. Santoso, "Perbandingan Performa SVM dan Naïve Bayes Pada Analisis Sentimen Aplikasi Game Online," vol. 8, no. 1, pp. 21-30, 2024.