DIGIVENTURE
Journal of Applied Digital Business Research

Vol. 2, No. 1, (2026), pp. 137-160
e-ISSN: 3123-5212
Website: https://journal.unesa.ac.id/digiventure

Comparative Analysis of TD-LSTM And GRU for Aspect-Based Sentiment Analysis of Gamification Elements in Shopee User Reviews

Annisa Esadora Hartanto[1] and Anita Safitri[2]

[1,2] Digital Business, Universitas Negeri Surabaya, Indonesia

*) Corresponding Author (Email: annisaessadorahartanto@gmail.com)

ABSTRACT

Gamification has become an important strategy for enhancing user engagement in e-commerce platforms. However, users may express different sentiments toward individual gamification elements, requiring a more detailed analysis of their experiences. This study aims to compare the performance of Target-Dependent Long Short-Term Memory (TD-LSTM) and Gated Recurrent Unit (GRU) in classifying aspect-level sentiment toward gamification elements in Shopee user reviews. Indonesian-language reviews were collected from the Google Play Store for the period January 2021 to December 2025. Following data cleaning, gamification-related filtering, and aspect identification based on the Werbach and Hunter framework, 7,964 review–aspect pairs were constructed and manually annotated into positive, neutral, and negative sentiment categories. TD-LSTM was employed as the baseline model, while a bidirectional GRU architecture served as the comparative model. Both models were evaluated using accuracy, precision, recall, and F1-score. The results showed that GRU achieved an accuracy of 88.71% and a weighted F1-score of 88.22%, outperforming TD-LSTM, which achieved 74.78% accuracy and a reported F1-score of 76.76%. These findings indicate that the implemented GRU model provides stronger classification performance for aspect-based sentiment analysis of Indonesian e-commerce reviews. The study contributes to the evaluation of deep learning approaches for analyzing user perceptions of gamification features in digital commerce.

Keywords: *Aspect-Based Sentiment Analysis, Gamification, Shopee, TD-LSTM, Gated Recurrent Unit*

Copyright © 2026 by Authors; this is an open-access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0).

I. INTRODUCTION

The rapid development of the e-commerce industry has intensified competition among digital marketplaces, encouraging businesses to implement innovative strategies to attract users and maintain long-term engagement. In Indonesia, the e-commerce sector continues to experience substantial growth, with increasing competition among major platforms such as Shopee and Tokopedia (Suratmanto et al., 2025). One strategy increasingly adopted by digital platforms is gamification, which refers to the integration of game design elements into non-game contexts to enhance user motivation, participation, and engagement (Deterding et al., 2011; Werbach & Hunter, 2012). Gamification mechanisms, including points, rewards, challenges, missions, leaderboards, and progress systems, have been widely implemented to encourage repeated interactions and create more engaging digital experiences (Putra Rahmadhan et al., 2023; Stieglitz et al., 2017). Furthermore, previous studies have demonstrated that gamification can enhance users' hedonic experiences, strengthen emotional attachment, and contribute to sustained engagement with digital platforms (Bitrián et al., 2021; Dwijaputra et al., 2025) .

As one of the prominent e-commerce platforms in Indonesia, Shopee incorporates various gamification mechanisms into its mobile application to enhance user engagement and encourage continuous platform interaction. These mechanisms include Shopee Coins, daily check-ins, Shopee Games, lucky draws, promotional challenges, and reward-based activities. Through these features, Shopee integrates entertainment-oriented experiences with conventional shopping activities, providing users with additional incentives to interact with the platform (Bulu & Natalia, 2024) . Previous research has highlighted the role of gamification in enhancing user engagement and strengthening relationships between consumers and digital platforms (Hollebeek et al., 2021) . Nevertheless, the implementation of gamification does not necessarily generate uniformly positive experiences, as users may perceive individual game elements differently depending on their expectations, experiences, and the benefits obtained. Therefore, understanding user perceptions of specific gamification elements is essential for evaluating their implementation and identifying opportunities for improvement.

Online customer reviews provide a valuable source of user-generated information for examining consumer perceptions of digital services. Reviews published on application distribution platforms, such as the Google Play Store, contain opinions, experiences, complaints, and evaluations that reflect users' interactions with mobile applications. These reviews can provide meaningful insights into user satisfaction, perceived service quality, and application functionality, allowing businesses to identify recurring concerns and potential improvements (Dąbrowski et al., 2022; Dhar & Bose, 2022) . However, the substantial volume of user-generated reviews makes manual evaluation increasingly

time-consuming and difficult to implement consistently. Consequently, computational approaches based on Natural Language Processing (NLP) have become important tools for extracting and interpreting information from unstructured textual data (Malik & Bilal, 2023; Sumarjan et al., 2024) .

Sentiment analysis, also referred to as opinion mining, is a computational approach used to identify and classify opinions, attitudes, and emotional evaluations expressed in textual information (Liu, 2012). Conventional sentiment analysis generally determines the overall polarity of a document or sentence, classifying it as positive, negative, or neutral. Although this approach provides an overview of user perceptions, it has limitations when applied to reviews containing multiple opinions about different application features. For instance, a user may express satisfaction with rewards received through Shopee Coins while simultaneously criticizing the complexity of completing daily missions. Document-level sentiment classification may fail to distinguish these contrasting evaluations because it assigns a single sentiment orientation to the entire review. To address this limitation, Aspect-Based Sentiment Analysis (ABSA) enables the identification of specific aspects discussed within a text and determines the sentiment associated with each aspect individually (Hu & Liu, 2004; Liu, 2012). This approach provides a more detailed understanding of user perceptions and supports feature-specific evaluations of digital applications.
The application of ABSA has increasingly incorporated deep learning techniques to capture contextual relationships between aspect terms and sentiment expressions. Among the neural architectures developed for aspect-level sentiment classification, Target-Dependent Long Short-Term Memory (TD-LSTM) represents an established approach that explicitly considers the contextual information surrounding a target aspect. Introduced by (Tang et al., 2016) TD-LSTM processes the left and right contexts of a target term to generate representations for sentiment prediction. By incorporating target-dependent contextual information, the model provides a foundation for identifying sentiment polarity toward specific aspects within user-generated reviews. Nevertheless, the effectiveness of recurrent architectures depends on their ability to capture relevant contextual information and manage sequential dependencies in textual data.

Another recurrent neural network architecture widely applied to sequential data processing is the Gated Recurrent Unit (GRU) (Cho et al., 2014). Unlike conventional LSTM architectures, GRU employs a simplified gating mechanism consisting of update and reset gates to regulate information flow and capture dependencies within sequential inputs. This architecture offers an alternative approach to contextual representation learning while employing a relatively compact recurrent structure. In sentiment classification tasks, GRU-based architectures can learn contextual patterns associated with positive, negative, and neutral expressions. However, their effectiveness in aspect-level classification requires empirical evaluation within specific linguistic and application contexts. Accordingly, comparing TD-LSTM and GRU provides an opportunity to examine the

classification performance of two recurrent architectures with different contextual processing mechanisms when applied to Indonesian-language e-commerce reviews.

Despite the growing body of research on gamification and sentiment analysis, the integration of aspect-level sentiment classification with a structured gamification framework remains an important area for further investigation. Previous studies have examined gamification primarily in relation to user engagement, motivation, satisfaction, and behavioral intentions (Bitrián et al., 2021; Bulu & Natalia, 2024; Xi & Hamari, 2019). Meanwhile, ABSA research has demonstrated the usefulness of neural architectures for identifying sentiment toward specific targets within textual data (Tang et al., 2016; Wang et al., 2016). However, these research streams do not fully address how individual gamification elements implemented in Indonesian e-commerce applications are perceived through naturally occurring user reviews. In particular, empirical investigation of aspect-level sentiment toward Shopee's gamification elements using the Werbach and Hunter framework, alongside a focused comparison of TD-LSTM and GRU, remains the specific research problem addressed in this study.

To address this research problem, the present study applies ABSA to Indonesian-language user reviews of the Shopee application collected from the Google Play Store. Gamification aspects are identified using a domain-specific dictionary developed with reference to the gamification framework proposed by Werbach and Hunter (Werbach & Hunter, 2012) which categorizes game elements into dynamics, mechanics, and components. The identified aspects are subsequently transformed into review–aspect pairs and manually annotated according to three sentiment categories: positive, neutral, and negative. TD-LSTM is employed as the baseline model, while GRU serves as the comparative model for evaluating aspect-level sentiment classification performance. Both models are assessed using accuracy, precision, recall, and F1-score to examine their effectiveness in classifying user perceptions of gamification elements.

This study aims to compare the classification performance of TD-LSTM and GRU in aspect-based sentiment analysis, identify the distribution of user sentiment toward gamification elements implemented in Shopee, and determine which elements receive predominantly positive or negative evaluations. The study contributes to the literature by integrating a theoretically grounded gamification framework with deep learning-based sentiment classification in the context of Indonesian e-commerce reviews. From a methodological perspective, it provides empirical evidence regarding the performance of two recurrent neural network architectures for aspect-level sentiment classification. From a practical perspective, the findings offer data-driven insights that may assist e-commerce platform developers and digital business practitioners in evaluating gamification features, identifying sources of user dissatisfaction, and prioritizing improvements to the user experience.

II. LITERATURE REVIEW

Gamification in E-Commerce Platforms

Gamification refers to the application of game design elements in non-game contexts to enhance user motivation, engagement, and participation (Deterding et al., 2011). Unlike conventional games, which are primarily designed for entertainment, gamification integrates selected game mechanisms into existing systems to encourage specific behaviors and improve user experiences. In digital business environments, gamification has become an increasingly relevant strategy for encouraging repeated interactions, strengthening user engagement, and supporting customer retention (Stieglitz et al., 2017). Previous studies have demonstrated that gamification can generate both hedonic and utilitarian value by combining enjoyable experiences with tangible incentives, including points, discounts, and rewards (Hollebeek et al., 2021).

Werbach and Hunter (2012) proposed a structured gamification framework consisting of three hierarchical categories: dynamics, mechanics, and components. Dynamics represent the highest-level elements that shape user experiences, including emotions, progression, constraints, narratives, and relationships. Mechanics describe the processes that encourage user participation, such as challenges, competition, cooperation, feedback, and rewards. Components represent the concrete manifestations of gamification, including points, badges, achievements, leaderboards, and levels. This framework provides a systematic foundation for identifying and evaluating gamification elements within digital platforms.

In e-commerce applications, gamification is commonly implemented through reward systems, daily check-ins, interactive challenges, and promotional games to encourage users to engage beyond conventional purchasing activities (Putra Rahmadhan et al., 2023). Shopee incorporates these mechanisms through features such as Shopee Coins, Shopee Games, daily login rewards, and promotional challenges. However, the effectiveness of gamification should not be evaluated solely through the availability of game elements, as users may express different perceptions of individual features. Therefore, the Werbach and Hunter framework is adopted in this study to operationalize gamification elements into 30 analytical aspects, allowing user perceptions to be examined systematically through application reviews.

Aspect-Based Sentiment Analysis

Sentiment analysis is a Natural Language Processing (NLP) technique used to identify and classify subjective opinions expressed in textual data. According to Liu (2012), sentiment analysis can be performed at different analytical levels, including document-level, sentence-level, and aspect-level classification. Conventional document-level sentiment analysis assigns an overall sentiment polarity to an entire text, whereas sentence-level analysis evaluates sentiment

within individual sentences. However, these approaches may overlook differences in sentiment when users discuss multiple aspects within the same review.

Aspect-Based Sentiment Analysis (ABSA) addresses this limitation by identifying specific opinion targets and determining the sentiment associated with each target (Hu & Liu, 2004). For example, an application review may express positive sentiment toward a reward system while simultaneously criticizing the difficulty of completing daily missions. ABSA enables these contrasting opinions to be analyzed independently, producing more detailed information about user perceptions.

The development of ABSA has been supported by standardized evaluation initiatives, including SemEval, which established benchmark tasks and datasets for aspect-level sentiment classification (Pontiki et al., 2016). More recently, deep learning approaches have been employed to capture contextual dependencies between aspect terms and surrounding sentiment expressions. These approaches are particularly relevant to user-generated reviews, where informal language, contextual ambiguity, and multiple opinion targets may complicate sentiment classification.

In the context of Shopee, ABSA provides a suitable approach for examining how users perceive individual gamification elements rather than evaluating overall application satisfaction. By associating each identified gamification aspect with a positive, neutral, or negative sentiment label, the analysis enables more specific evaluations of game-related features and their perceived user experiences.

Target-Dependent Long Short-Term Memory (TD-LSTM)

Target-Dependent Long Short-Term Memory (TD-LSTM) is a recurrent neural network architecture introduced by Tang et al. (2016) for target-dependent sentiment classification. The architecture extends conventional Long Short-Term Memory (LSTM) networks by explicitly incorporating contextual information surrounding a target term. Unlike conventional sequence classification models that represent an entire sentence uniformly, TD-LSTM separates the textual sequence into left and right contexts relative to the target aspect.

The left and right contextual sequences are processed through separate LSTM networks, and their final hidden representations are concatenated to construct a target-dependent representation for sentiment classification (Tang et al., 2016). This mechanism enables the model to distinguish sentiment expressions associated with a particular target from opinions referring to other aspects within the same text.

TD-LSTM is particularly relevant to ABSA because sentiment polarity is not always determined by the overall emotional orientation of a review. Instead, sentiment may depend on the relationship between a specific aspect and its surrounding contextual expressions. Previous studies have employed target-dependent recurrent

architectures to establish baseline performance and evaluate improvements in contextual sentiment modeling (Ma et al., 2017; Tang et al., 2016)

In this study, TD-LSTM serves as the baseline architecture because it explicitly incorporates aspect-related contextual information and provides an established reference for evaluating aspect-level sentiment classification. Its implementation enables the assessment of how target-dependent sequence representations perform when applied to Indonesian-language reviews discussing gamification elements in Shopee.

Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is a recurrent neural network architecture introduced by Cho et al.(2014) to capture sequential dependencies through a simplified gating mechanism. GRU was developed as an alternative to conventional recurrent architectures and employs two principal gates, namely the update gate and reset gate, to regulate the flow of information across sequential inputs.

The update gate determines the extent to which previously learned information is retained, whereas the reset gate controls the contribution of previous information when generating a new candidate hidden state. Unlike LSTM, which maintains a separate memory cell and employs input, forget, and output gates, GRU integrates information regulation into a more compact architecture. Previous research has demonstrated that gated recurrent architectures can effectively capture sequential dependencies in natural language processing tasks while offering a different balance between model complexity and predictive performance (Cho et al., 2014; Chung et al., 2014).

In sentiment analysis, GRU can learn contextual representations from textual sequences and identify linguistic patterns associated with different sentiment categories. Its ability to retain relevant sequential information makes it applicable to user reviews containing informal expressions and context-dependent opinions. Nevertheless, GRU does not inherently employ the same explicit left–right target separation as TD-LSTM. Consequently, the two architectures provide different approaches to representing contextual information for sentiment classification.

In the present study, a bidirectional GRU architecture is implemented as the comparative model to evaluate its ability to classify sentiment toward predefined gamification aspects. The comparison with TD-LSTM is intended to examine differences in classification performance between explicit target-dependent contextual modeling and gated sequential representation learning within the same application review dataset.

Previous Studies and Research Gap

Previous research has established the theoretical and methodological foundations for examining gamification and aspect-level sentiment classification. Deterding et al. (2011) and Werbach and Hunter (2012) provided foundational

definitions and frameworks for understanding gamification, while Hamari et al. (2014) reviewed empirical evidence concerning the effectiveness of gamification in influencing user motivation and engagement. In the context of digital platforms, Xi and Hamari (2019) investigated the relationship between gamification features and users' psychological need satisfaction, emphasizing the importance of examining different gamification mechanisms rather than treating gamification as a single uniform construct.

In sentiment analysis research, Liu (2012) established the conceptual foundation for opinion mining, while Pontiki et al. (2016) contributed standardized ABSA evaluation tasks and datasets. Tang et al. (2016) introduced target-dependent recurrent architectures for sentiment classification, demonstrating the relevance of incorporating target-specific contextual information. Meanwhile, Cho et al. (2014) and Chung et al. (2014) established the theoretical and empirical foundations of GRU-based sequential learning, supporting its application as an alternative recurrent architecture for textual classification.

Although these studies provide important insights into gamification and sentiment classification, their primary contributions address different research objectives. Gamification studies generally emphasize engagement, motivation, and behavioral outcomes, whereas neural sentiment classification studies focus on linguistic representation and predictive performance. The application of these approaches to evaluating individual gamification elements through Indonesian-language e-commerce reviews therefore warrants further investigation.

Accordingly, this study integrates the gamification framework proposed by Werbach and Hunter (2012) with ABSA to examine user perceptions of Shopee's gamification features. Unlike studies that evaluate overall application sentiment, this research identifies individual gamification aspects and classifies their associated sentiment polarity. Furthermore, the study compares TD-LSTM and GRU under a shared dataset and evaluation framework to investigate their performance in aspect-level sentiment classification. This approach is expected to contribute empirical evidence regarding recurrent neural network performance while providing practical insights into user perceptions of gamification features in Indonesian e-commerce applications.

III. METHOD

Research Design

This study employs a quantitative research approach using Aspect-Based Sentiment Analysis (ABSA) to investigate user perceptions of gamification elements implemented in the Shopee mobile application. The research integrates Natural Language Processing (NLP), dictionary-based aspect identification, manual sentiment annotation, and deep learning-based classification. Two recurrent neural

network architectures are evaluated: Target-Dependent Long Short-Term Memory (TD-LSTM) as the baseline model and Gated Recurrent Unit (GRU) as the comparative model.

The research process consists of seven stages: data collection, text preprocessing, gamification-related review filtering, aspect identification, review–aspect pair construction, sentiment annotation, and model training and evaluation. The gamification elements are identified based on the framework proposed by Werbach and Hunter (2012), while sentiment classification is conducted using supervised deep learning. The resulting classifications are subsequently used to examine the distribution of user perceptions across individual gamification aspects.

Data Collection and Dataset Construction

The dataset was obtained from user reviews of the Shopee application available on the Indonesian Google Play Store (package name: com.shopee.id). Data collection was conducted using the google_play_scraper Python library, with the language and country parameters configured as lang='id' and country='id', respectively. Reviews were retrieved using the newest-first sorting mechanism and filtered according to the observation period from January 2021 to December 2025.

The initial collection produced 2,994,050 user reviews. Following duplicate removal, 1,743,710 unique reviews were retained for subsequent processing. Since this study focuses specifically on gamification-related experiences, additional filtering was performed to exclude reviews unrelated to gamification. This process identified 8,015 candidate reviews. Following aspect identification, review–aspect pair construction, and annotation, the final dataset consisted of 7,964 review–aspect pairs.

Table 1. Research Dataset Construction

Dataset Construction Stage	Number of Data
Raw reviews collected	2,994,050
Reviews after duplicate removal	1,743,710
Candidate gamification reviews	8,015
Final review–aspect pairs	7,964
Training set	6,368
Validation set	796
Test set	797

Source: Author's research data.

The final dataset was partitioned into training, validation, and testing subsets using an approximately 80:10:10 allocation with stratification according to sentiment labels. The training subset was used for parameter optimization, the validation subset for monitoring model performance during training, and the testing subset for final classification evaluation.

Text Preprocessing and Gamification Filtering

Text preprocessing was performed to improve the consistency and quality of user-generated reviews before aspect identification and sentiment classification. The preprocessing procedures included text cleaning, case normalization, removal of irrelevant symbols and URLs, normalization of informal expressions, stopword processing, and text validation. Indonesian-language processing was supported by the Sastrawi library, while negation expressions such as "tidak," "gak," "ga," and "bukan" were preserved to avoid altering the original sentiment polarity.

Following preprocessing, anchor-based filtering was implemented to identify reviews potentially discussing gamification. Three keyword groups were employed: strong anchors, weak anchors, and exclusion keywords. Strong anchors included explicit gamification-related expressions such as "Shopee Games," "koin," "spin," "misi," "check-in," and "reward." Reviews containing strong anchors were retained as potential gamification-related observations. Weak anchors were used to identify less explicit references, with reviews retained when at least two distinct weak-anchor expressions were present. Exclusion keywords were applied to remove discussions primarily concerning unrelated application services, including logistics, payment services, product delivery, and seller performance.

This filtering procedure was designed to reduce irrelevant observations while retaining textual information associated with users' experiences of gamification features.

Gamification Aspect Identification and Sentiment Annotation

Gamification aspects were identified using a domain-specific dictionary developed based on the framework proposed by Werbach and Hunter (2012). The framework classifies gamification elements into three categories: dynamics, mechanics, and components. In this study, 30 gamification elements were operationalized as analytical aspects and mapped to corresponding expressions found in Shopee user reviews.

Aspect identification was conducted through dictionary-based keyword matching and contextual filtering. Each detected aspect was associated with its corresponding review to construct an individual review–aspect pair. Consequently, a review discussing multiple gamification elements could generate more than one observation, allowing sentiment to be classified separately for each aspect.

The resulting review–aspect pairs were manually annotated into three sentiment categories: positive, neutral, and negative. Positive sentiment represented favorable evaluations of a gamification element, negative sentiment indicated dissatisfaction or criticism, and neutral sentiment represented expressions without a clearly favorable or unfavorable evaluation. Annotation was conducted by a single annotator following predefined labeling guidelines. The annotated dataset served as the ground-truth data for supervised model training and evaluation.

Deep Learning Model Development

Two recurrent neural network architectures were implemented to classify aspect-level sentiment. TD-LSTM was selected as the baseline because it explicitly

models contextual information surrounding the target aspect, while GRU was selected as the comparative architecture because of its gated sequential learning mechanism.

Target-Dependent Long Short-Term Memory (TD-LSTM)

TD-LSTM, introduced by Tang et al. (2016), processes contextual information located before and after a target aspect through separate LSTM networks. In this study, the review text was divided into left and right contextual sequences relative to the identified gamification aspect. Each contextual sequence was transformed into numerical representations through a trainable embedding layer with 128 dimensions.

The left and right contextual sequences were processed using separate LSTM layers containing 128 hidden units, with the right context encoded in the backward direction. The resulting contextual representations were concatenated and passed through a fully connected layer containing 64 neurons with Rectified Linear Unit (ReLU) activation. A final Softmax classification layer generated probability distributions across the three sentiment categories.

The TD-LSTM model was trained using the Adam optimizer with a learning rate of 0.001 and sparse categorical crossentropy as the loss function. Training employed a batch size of 32 and a maximum of 15 epochs. Early stopping was implemented by monitoring validation loss with a patience value of three epochs. Balanced class weights were applied to reduce the influence of unequal sentiment class distributions.

Gated Recurrent Unit (GRU)

GRU, introduced by Cho et al. (2014), is a recurrent neural network architecture that regulates sequential information through update and reset gates. In this study, a bidirectional GRU architecture was employed to capture contextual information from both forward and backward directions.

Unlike TD-LSTM, which separately processes contextual sequences surrounding the target aspect, the GRU model received a single input sequence constructed by combining the identified aspect and its corresponding review text. The input sequence was tokenized and transformed using a trainable embedding layer with 300 dimensions, followed by a bidirectional GRU layer containing 128 hidden units.

To reduce overfitting, a dropout layer with a rate of 0.5 was applied after the recurrent layer. The resulting representation was subsequently processed through a fully connected layer containing 64 neurons with ReLU activation, followed by an additional dropout layer with a rate of 0.3. The final Softmax layer classified each observation into positive, neutral, or negative sentiment.

The GRU model was trained using the Adam optimizer with a learning rate of 0.001 and sparse categorical crossentropy as the loss function. The training configuration included a batch size of 32, a maximum of 15 epochs, and early stopping based on validation loss with a patience value of three epochs.

Table 2. Model Configuration

Parameter	TD-LSTM	GRU
Input representation	Left and right contexts	Aspect and review text
Embedding dimension	128	300
Recurrent layer	Two LSTM branches	Bidirectional GRU
Hidden units	128 per branch	128
Dense layer	64 (ReLU)	64 (ReLU)
Output layer	Softmax	Softmax
Optimizer	Adam	Adam
Learning rate	0.001	0.001
Batch size	32	32
Maximum epochs	15	15
Early stopping patience	3	3
Loss function	Sparse categorical crossentropy	Sparse categorical crossentropy

Source: Author's experimental configuration.

Model Evaluation

Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrices. Accuracy measures the proportion of correctly classified observations, while precision evaluates the proportion of predicted instances belonging to a sentiment class that are correctly classified. Recall measures the proportion of actual instances correctly identified by the model, whereas the F1-score represents the harmonic mean of precision and recall.

In addition to aggregate evaluation metrics, confusion matrices were examined to identify classification errors across positive, neutral, and negative sentiment categories. The comparative evaluation was conducted using the held-out testing dataset to assess the predictive performance of TD-LSTM and GRU.

Aspect-Level Sentiment Analysis

Following sentiment classification, aspect-level analysis was conducted to examine user perceptions of gamification elements implemented in Shopee. Sentiment distributions were summarized according to the identified gamification aspects and their corresponding positive, neutral, and negative labels.

The analysis focused on identifying gamification elements associated with dominant positive and negative evaluations, as well as examining differences in sentiment distributions across the dynamics, mechanics, and components categories proposed by Werbach and Hunter (2012). Descriptive analysis was used to interpret the resulting sentiment patterns and identify gamification features that may require further evaluation or improvement.

This approach combines predictive sentiment classification with a theoretically grounded interpretation of gamification elements, allowing the findings to contribute to both computational sentiment analysis and practical digital business evaluation.

IV. RESULTS AND DISCUSSION

Dataset Preparation and Characteristics

The dataset preparation process transformed 2,994,050 Google Play Store reviews into a structured corpus for aspect-based sentiment classification. Following duplicate removal, 1,743,710 unique reviews were retained. Gamification-related filtering subsequently identified 8,015 candidate reviews, which underwent aspect identification, review–aspect pair construction, and manual sentiment annotation. This process yielded 7,964 review–aspect pairs associated with gamification elements in the Shopee application.

The experimental datasets used for model development and evaluation comprised 7,961 review–aspect pairs, consisting of 6,368 training observations, 796 validation observations, and 797 testing observations. Both TD-LSTM and GRU were evaluated using the same testing dataset to support a consistent comparison of their predictive performance.

The dataset exhibited an imbalanced distribution across the three sentiment categories. Negative sentiment represented the largest proportion, followed by positive sentiment, while neutral observations were substantially less frequent. In the testing dataset, 471 observations were labeled negative, 296 positive, and only 30 neutral. This imbalance is particularly relevant to the evaluation of class-specific performance, as the limited representation of neutral observations may restrict the models' ability to recognize this category reliably.

The substantial reduction from the initial review collection to the final gamification-related corpus highlights the importance of domain-specific filtering. Most general application reviews concerned topics outside the research scope, including logistics, payment services, and product transactions. The combination of anchor-based filtering and dictionary-based aspect identification therefore enabled the analysis to focus on user experiences associated with individual gamification elements.

These dataset characteristics provide important context for interpreting the classification results, particularly the differences between aggregate performance metrics and the recognition of underrepresented sentiment categories.

Performance of TD-LSTM

The TD-LSTM model was implemented as the baseline architecture for aspect-level sentiment classification. The model separately processed contextual information surrounding each identified gamification aspect before combining the resulting representations for sentiment prediction.

The experimental results indicate that TD-LSTM achieved an accuracy of 74.78%, precision of 81.02%, recall of 74.78%, and F1-score of 76.76% on the testing dataset. These results demonstrate that target-dependent contextual representations can identify sentiment patterns associated with individual gamification elements.

Nevertheless, the difference between precision and recall suggests that the model's predictive performance was not equally strong across all observations. Although TD-LSTM explicitly incorporated contextual information surrounding target aspects, its performance remained limited when handling ambiguous sentiment expressions and unequal sentiment class distributions.

The training results also indicate that TD-LSTM was capable of learning patterns from the annotated reviews. However, the lower testing performance suggests that successful optimization during training does not necessarily guarantee equally strong generalization to previously unseen review–aspect pairs.

Performance of GRU

The bidirectional GRU model was implemented as the comparative architecture for aspect-level sentiment classification. Unlike TD-LSTM, which processes left and right contextual sequences separately, the GRU model combined the target aspect and its corresponding review text into a single input sequence to capture contextual information in both directions.

Based on the experimental evaluation using 797 testing observations, the GRU model achieved an accuracy of 88.71%, weighted precision of 88.14%, weighted recall of 88.71%, and weighted F1-score of 88.22%. These results demonstrate that the implemented GRU model achieved strong classification performance on the held-out testing dataset and outperformed the TD-LSTM baseline across all four evaluation metrics.

The bidirectional recurrent architecture enabled the model to capture sequential information from both forward and backward contexts, while its update and reset gates regulated information retention. These mechanisms may have contributed to its ability to recognize sentiment patterns in Indonesian-language e-commerce reviews. Nevertheless, the observed performance should be interpreted in relation to the complete model configuration rather than attributed solely to the recurrent architecture.

Despite its relatively high aggregate performance, class-specific evaluation remains important, particularly for neutral sentiment, which was underrepresented in the annotated dataset. Therefore, the GRU model's overall performance should be considered alongside its classification report and confusion matrix to assess potential differences in predictive performance across sentiment categories.

Comparative Performance of TD-LSTM and GRU

The comparative evaluation was conducted using four classification metrics: accuracy, precision, recall, and F1-score. The results are presented in Table 3.

Table 3. Performance Comparison of TD-LSTM and GRU

Evaluation Metric	TD-LSTM	GRU
Accuracy	74.78%	88.71%
Precision	81.02%	88.14%
Recall	74.78%	88.71%
F1-score	76.76%	88.22%

Source: Author's experimental results

The comparative evaluation demonstrates that the bidirectional GRU model outperformed TD-LSTM across all four evaluation metrics, as presented in Table 3. GRU achieved an accuracy of 88.71%, weighted precision of 88.14%, weighted recall of 88.71%, and weighted F1-score of 88.22%. In comparison, TD-LSTM achieved an accuracy of 74.78%, weighted precision of 81.02%, weighted recall of 74.78%, and weighted F1-score of 76.76%. These results represent improvements of 13.93 percentage points in accuracy, 7.12 percentage points in weighted precision, 13.93 percentage points in weighted recall, and 11.46 percentage points in weighted F1-score.

The higher classification performance of GRU suggests that its implemented bidirectional architecture was more effective in capturing contextual information relevant to sentiment classification in Indonesian-language Shopee reviews. By processing the combined aspect and review text in both forward and backward directions, the model could learn contextual relationships associated with positive, neutral, and negative expressions. This finding is consistent with the capacity of gated recurrent architectures to model sequential dependencies, as discussed by Cho et al. (2014) and Chung et al. (2014).

Nevertheless, the observed performance differences cannot be attributed solely to the recurrent architectures. The two models differed in input representation, embedding dimensions, and training configurations, including the application of class weighting in TD-LSTM but not in GRU. Consequently, the findings should be interpreted as a comparison between the implemented modeling pipelines rather than as evidence of the inherent superiority of GRU over TD-LSTM under identical experimental conditions.

Overall, the results indicate that the implemented bidirectional GRU model provided stronger predictive performance for aspect-based sentiment classification of Shopee gamification reviews. Its performance also supports its use for the subsequent descriptive analysis of sentiment distributions across individual gamification elements.

Classification Error Analysis

Class-specific evaluation was conducted to examine differences in the predictive performance of TD-LSTM and GRU across negative, neutral, and positive sentiment categories. Although the bidirectional GRU model achieved higher overall classification performance, both models exhibited variations in their ability to identify individual sentiment classes.

The classification reports showed that GRU achieved F1-scores of 0.91, 0.40, and 0.88 for negative, neutral, and positive sentiment, respectively. In comparison, TD-LSTM obtained corresponding F1-scores of 0.80, 0.23, and 0.78. These results indicate that GRU provided stronger classification performance across all three sentiment categories, with particularly notable improvements in identifying negative and positive expressions.

Neutral sentiment remained the most challenging category for both models. The testing dataset contained only 30 neutral observations, compared with 471 negative and 296 positive observations. For neutral sentiment, TD-LSTM achieved a precision of 0.15 and recall of 0.43, whereas GRU achieved a precision of 0.60 and recall of 0.30. Although GRU produced more precise neutral predictions, its lower recall indicates that a substantial proportion of actual neutral observations remained unrecognized. This finding highlights the importance of examining class-specific metrics rather than relying exclusively on overall accuracy.

The unequal distribution of sentiment classes may have contributed to these classification difficulties. Furthermore, Indonesian-language user reviews frequently contain informal expressions, implicit evaluations, and context-dependent opinions that can complicate sentiment interpretation. These linguistic characteristics may introduce additional ambiguity, particularly when distinguishing neutral expressions from positive or negative evaluations.

Overall, the classification error analysis demonstrates that the GRU implementation improved sentiment recognition across the evaluated classes but did not fully resolve the challenges associated with minority-class classification. Future research could investigate alternative sampling strategies, class-balancing techniques, and more extensive annotation procedures to improve the recognition of underrepresented sentiment categories.

GRU-Based Aspect-Level Sentiment Mapping

Following the comparative model evaluation, a GRU-based aspect-level sentiment mapping was conducted to examine user perceptions of individual gamification elements in the Shopee application. The analysis employed predictions generated by the trained bidirectional GRU model for 7,961 review–aspect pairs. Unlike the held-out test evaluation, this mapping was performed on the combined experimental corpus to provide a descriptive overview of sentiment patterns across gamification aspects.

The overall sentiment distribution showed that negative sentiment accounted for 60.91% (4,849 pairs), followed by positive sentiment at 37.19% (2,961 pairs) and neutral sentiment at 1.90% (151 pairs). These results indicate that negative evaluations were more prevalent than positive or neutral evaluations among the review–aspect pairs analyzed by the GRU model.

Table 4. GRU-Based Sentiment Distribution Across Major Gamification Aspects

Gamification Aspect	Total Pairs	Negative (%)	Neutral (%)	Positive (%)
Points	4,658	58.52	1.76	39.72
Narrative	633	88.63	1.26	10.11
Rewards	617	47.00	5.35	47.65
Emotions	555	57.12	0.00	42.88
Resource Acquisition	542	34.32	0.00	65.68
Challenges	252	94.84	0.00	5.16
Win States	204	85.29	0.49	14.22

Source: Author's GRU-based sentiment mapping results.

As presented in Table 4, Points was the most frequently discussed gamification aspect, accounting for 4,658 review–aspect pairs. Negative sentiment represented 58.52% of these observations, while positive sentiment accounted for 39.72%. This finding suggests that the prominence of point-based mechanisms in user reviews does not necessarily indicate favorable perceptions. Instead, users expressed a greater proportion of negative evaluations toward these mechanisms.

Several gamification aspects exhibited particularly high proportions of negative sentiment. Challenges recorded the highest negative proportion among the aspects presented, at 94.84%, followed by Narrative at 88.63% and Win States at 85.29%. These distributions indicate that the GRU model identified predominantly unfavorable evaluations associated with these elements. However, the underlying reasons for dissatisfaction cannot be determined from sentiment proportions alone and would require further examination of the corresponding review content.

In contrast, Resource Acquisition exhibited a predominantly positive sentiment distribution, with 65.68% positive and 34.32% negative predictions. This pattern suggests that users may respond more favorably to gamification elements associated with obtaining resources or benefits. Rewards showed a comparatively balanced distribution, with 47.65% positive and 47.00% negative sentiment, indicating mixed user evaluations rather than a clear positive or negative tendency.

Figure 1. GRU-Based Sentiment Distribution Across Major Gamification Aspects



The visualization in Figure 1 illustrates differences in sentiment proportions across frequently discussed gamification aspects. Although several elements were predominantly associated with negative predictions, the distribution varied considerably between aspects, highlighting the importance of analyzing gamification features individually rather than treating gamification as a single, uniformly evaluated experience.

These findings are consistent with the perspective that gamification effectiveness depends on the value and experience created through user interactions rather than the mere presence of game design elements (Huotari & Hamari, 2017). From a digital business perspective, aspect-level sentiment mapping can help identify gamification mechanisms that warrant further investigation and those associated with relatively favorable user responses.

Nevertheless, these results should be interpreted as descriptive distributions of GRU-predicted sentiment within the experimental corpus. Because the mapping included training, validation, and testing observations, it does not constitute an independent evaluation of the model's generalization performance. Furthermore, the observed sentiment distributions represent the analyzed review–aspect pairs and should not be generalized directly to the entire Shopee user population.

Discussion and Implications

The findings demonstrate that the implemented bidirectional GRU model achieved stronger overall performance than the TD-LSTM baseline for aspect-based sentiment classification of Shopee gamification reviews. GRU obtained an accuracy

of 88.71% and a weighted F1-score of 88.22%, compared with 74.78% accuracy and a weighted F1-score of 76.76% for TD-LSTM. These results indicate that the GRU-based modeling pipeline was more effective in identifying sentiment patterns within the evaluated Indonesian-language review–aspect pairs.

From a methodological perspective, the results highlight the potential of bidirectional recurrent learning for aspect-based sentiment analysis. While TD-LSTM explicitly processes contextual information surrounding a target aspect, the GRU implementation integrates the aspect and review text into a single sequence and captures contextual relationships in both directions. Nevertheless, the performance differences should be interpreted cautiously because the models employed different input representations, embedding dimensions, and class-weighting configurations. Therefore, the findings demonstrate the relative performance of the implemented pipelines rather than the inherent superiority of one recurrent architecture under identical experimental conditions.

Beyond classification performance, the GRU-based aspect-level mapping revealed substantial differences in sentiment distributions across gamification elements. Negative sentiment accounted for 60.91% of the analyzed review–aspect pairs, indicating that unfavorable evaluations were prevalent within the experimental corpus. However, this pattern was not uniform across individual gamification elements.

Points emerged as the most frequently discussed aspect, with 4,658 review–aspect pairs, of which 58.52% were classified as negative. Challenges and Narrative exhibited particularly high negative sentiment proportions of 94.84% and 88.63%, respectively. These findings suggest that frequently discussed gamification features may warrant further investigation, particularly when their associated reviews contain predominantly unfavorable evaluations. Nevertheless, sentiment distributions alone cannot establish the specific causes of user dissatisfaction.

In contrast, Resource Acquisition showed a predominantly positive sentiment distribution of 65.68%, while Rewards exhibited relatively balanced positive and negative proportions of 47.65% and 47.00%, respectively. These differences suggest that user perceptions of gamification mechanisms vary across individual elements. This interpretation aligns with the view that effective gamification depends on meaningful user experiences and perceived value rather than simply incorporating game design elements (Huotari & Hamari, 2017).

From a digital business perspective, the findings demonstrate the potential of ABSA as a feature-level evaluation tool for e-commerce platforms. By identifying which gamification aspects receive predominantly positive or negative evaluations, platform developers can prioritize further investigation of specific features, assess potential areas for improvement, and complement conventional engagement metrics with user-generated feedback. For example, aspects associated with high negative sentiment may be examined through qualitative review analysis to identify recurring

concerns, while positively evaluated aspects may provide useful references for future gamification development.

Several limitations should be acknowledged. First, sentiment annotation was conducted by a single annotator, which may introduce subjective labeling decisions. Second, the dataset was restricted to Indonesian-language Google Play Store reviews and may not represent the broader Shopee user population. Third, the unequal distribution of sentiment classes, particularly the limited number of neutral observations, may influence classification performance. Fourth, the GRU-based aspect mapping was conducted on a combined corpus containing training, validation, and testing observations; therefore, its distributions should be interpreted descriptively rather than as an independent assessment of model generalization.

Overall, the study contributes to applied digital business research by demonstrating how deep learning-based ABSA can support both comparative model evaluation and the identification of sentiment patterns associated with individual gamification elements. The findings provide a foundation for using aspect-level user feedback to inform more targeted and evidence-based evaluations of gamification features in e-commerce applications.

V. CONCLUSION AND RECOMMENDATION

Conclusion

This study compared the performance of Target-Dependent Long Short-Term Memory (TD-LSTM) and a bidirectional Gated Recurrent Unit (GRU) for aspect-based sentiment classification of gamification elements in Indonesian-language Shopee user reviews. The research applied a structured process involving review collection, gamification-related filtering, aspect identification based on the Werbach and Hunter framework, manual sentiment annotation, and supervised deep learning classification. The experimental dataset comprised 7,961 review–aspect pairs, divided into training, validation, and testing subsets.

The experimental results demonstrated that the implemented bidirectional GRU model outperformed the TD-LSTM baseline across all four evaluation metrics. GRU achieved an accuracy of 88.71%, weighted precision of 88.14%, weighted recall of 88.71%, and weighted F1-score of 88.22%. In comparison, TD-LSTM obtained an accuracy of 74.78%, weighted precision of 81.02%, weighted recall of 74.78%, and weighted F1-score of 76.76%. These findings indicate that the implemented GRU pipeline provided stronger predictive performance for classifying aspect-level sentiment in the evaluated dataset. However, the differences should not be attributed solely to the recurrent architectures because the models employed different input representations, embedding dimensions, and training configurations.

The GRU-based aspect-level sentiment mapping further revealed that negative sentiment dominated the analyzed corpus, accounting for 60.91% of the predicted review–aspect pairs, followed by positive sentiment at 37.19% and neutral sentiment at 1.90%. Points emerged as the most frequently discussed gamification aspect, with 4,658 observations, of which 58.52% were classified as negative. Challenges and Narrative exhibited particularly high proportions of negative sentiment, whereas Resource Acquisition was predominantly positive. Rewards showed a relatively balanced distribution of positive and negative sentiment. These findings demonstrate that user evaluations varied across gamification elements and that frequently discussed features were not necessarily associated with favorable sentiment.

Despite the stronger aggregate performance of GRU, both models experienced difficulties in recognizing neutral sentiment, which was substantially underrepresented in the testing dataset. Furthermore, the aspect-level mapping was conducted on a combined corpus containing training, validation, and testing observations. Accordingly, its results should be interpreted as descriptive distributions of model-predicted sentiment rather than an independent assessment of generalization performance.

Overall, the study demonstrates the potential of deep learning-based aspect-based sentiment analysis to support feature-level evaluations of gamification in e-commerce applications. By identifying sentiment patterns associated with individual gamification elements, the proposed analytical approach provides insights that may assist digital platforms in understanding user feedback and prioritizing further investigation of specific gamification features.

Research Limitations

Several limitations should be considered when interpreting the findings. First, the dataset was collected exclusively from Shopee user reviews published on the Google Play Store. Consequently, the findings may not represent perceptions expressed by users on other platforms, including the Apple App Store, social media, and alternative e-commerce applications.

Second, gamification aspect identification relied on a predefined dictionary and contextual filtering rules developed with reference to the Werbach and Hunter (2012) framework. Although this approach facilitated the identification of relevant review–aspect pairs, it may have excluded implicit expressions or previously unidentified terminology associated with gamification experiences.

Third, sentiment annotation was performed by a single annotator using predefined labeling guidelines. Since no inter-annotator agreement assessment was conducted, the reliability of the sentiment labels could not be independently established through measures such as Cohen's Kappa. Furthermore, the unequal distribution of sentiment categories, particularly the limited number of neutral observations, presented additional challenges for model training and evaluation.

Fourth, the comparison between TD-LSTM and GRU was based on their respective implementation configurations. Although both models were evaluated using the same research dataset and primary evaluation metrics, differences in input representation, embedding dimensions, and training strategies may have influenced their relative performance. Consequently, the study does not establish whether the GRU recurrent mechanism alone is responsible for the observed performance improvement.

Finally, the study focused on sentiment expressed in user-generated reviews rather than directly measuring user satisfaction, continued engagement, or purchasing behavior. Therefore, the findings should not be interpreted as evidence of causal relationships between gamification elements and behavioral outcomes.

Recommendations

Based on the findings, several recommendations are proposed for future research and practical application. First, subsequent studies should consider more balanced sentiment datasets, particularly by increasing the representation of neutral observations. Additional annotation procedures involving multiple annotators and inter-annotator agreement assessment could also improve the reliability of sentiment labels.

Second, future research should conduct more controlled architectural comparisons by standardizing input representations, embedding configurations, class-weighting strategies, and other relevant hyperparameters. Such experimental designs would help distinguish the effects of recurrent architecture choices from differences introduced by the broader modeling pipelines. Alternative approaches, including transformer-based models and other context-sensitive architectures, may also be evaluated to investigate potential improvements in minority-class classification.

Third, future aspect-level analyses should consider using independent evaluation datasets or cross-validation procedures to assess the stability and generalizability of sentiment distributions. Qualitative examination of representative user reviews is also recommended to identify the underlying reasons for positive and negative evaluations, as sentiment proportions alone cannot explain the specific experiences motivating user feedback.

From a practical perspective, e-commerce platform developers may use aspect-level sentiment analysis as a complementary tool for evaluating gamification features. Elements associated with predominantly negative sentiment, such as Challenges and Narrative, may be prioritized for further investigation, while relatively favorable responses toward Resource Acquisition may provide useful directions for feature evaluation. These insights should be considered alongside other evidence, including user behavior, engagement indicators, and qualitative feedback, before implementing design changes.

Finally, future studies may expand the dataset to additional review periods, e-commerce platforms, and linguistic contexts to examine whether similar sentiment patterns and model performance differences occur beyond the scope of the present research.

REFERENCES

Bitrián, P., Buil, I., & Catalán, S. (2021). Enhancing user engagement: The role of gamification in mobile apps. *Journal of Business Research*, *132*, 170–185. https://doi.org/10.1016/j.jbusres.2021.04.028

Bulu, D. S., & Natalia, E. C. (2024). The Effect of Gamification in Shopee Games on User's Brand Loyalty. *Mediator: Jurnal Komunikasi*, *17*(2), 257–269. https://doi.org/10.29313/mediator.v17i2.3039

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation*. http://arxiv.org/abs/1406.1078

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. http://arxiv.org/abs/1412.3555

Dąbrowski, J., Letier, E., Perini, A., & Susi, A. (2022). Analysing app reviews for software engineering: a systematic literature review. *Empirical Software Engineering*, *27*(2). https://doi.org/10.1007/s10664-021-10065-7

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). *CHI and apos;11 Extended Abstracts on Human Factors in Computing Systems*. Association for Computing Machinery.

Dhar, S., & Bose, I. (2022). Walking on air or hopping mad? Understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps. *Decision Support Systems*, *162*. https://doi.org/10.1016/j.dss.2022.113769

Dwijaputra, J., Purwitasari, D., & Zulaikha, E. (2025). How gamification influences e-commerce behavior in hardcore vs. casual gamers? *Telematics and Informatics Reports*, *18*. https://doi.org/10.1016/j.teler.2025.100212

Hollebeek, L. D., Das, K., & Shukla, Y. (2021). Game on! How gamified loyalty programs boost customer engagement value. *International Journal of Information Management*, *61*. https://doi.org/10.1016/j.ijinfomgt.2021.102308

Hu, M., & Liu, B. (2004). *Mining and Summarizing Customer Reviews*.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining 2*. https://doi.org/https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). *Interactive Attention Networks for Aspect-Level Sentiment Classification*. http://arxiv.org/abs/1709.00893

Malik, N., & Bilal, M. (2023). *Natural Language Processing for Analyzing Online Customer Reviews: A Survey, Taxonomy, and Open Research Challenges*. https://doi.org/10.20944/preprints202312.2210.v1

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., María Jiménez-Zafra, S., & Eryiğit, G. (2016). *SemEval-2016 Task 5: Aspect Based Sentiment Analysis*. http://alt.qcri.org/semeval2014/task4/

Putra Rahmadhan, M. A. W., Sensuse, D. I., Suryono, R. R., & Kautsarina. (2023). Trends and Applications of Gamification in E-Commerce: A Systematic Literature Review. *Journal of Information Systems Engineering and Business Intelligence*, *9*(1), 28–37. https://doi.org/10.20473/jisebi.9.1.28-37

Stieglitz, S., Lattemann, C., Robra-Bissantz, S., Zarnekow, R., & Brockmann, T. (2017). *Gamification* (S. Stieglitz, C. Lattemann, S. Robra-Bissantz, R. Zarnekow, & T. Brockmann, Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-319-45557-0

Sumarjan, N., Sulong, S. N., & Azeman, A. S. (2024). How online hotel reviews affect consumer booking decisions. In *Journal of Tourism, Hospitality & Culinary Arts (JTHCA) 2024* (Vol. 16, Number 2).

Suratmanto, B., Handarkho, Y. D., & Emanuel, A. W. R. (2025). The Influence of Gamification on Repurchase Intention at E-Marketplace from a Habit Perspective. *Journal of Information and Organizational Sciences*, *49*(1), 53–67. https://doi.org/10.31341/jios.49.1.4

Tang, D., Qin, B., Feng, X., & Liu, T. (2016). *Effective LSTMs for Target-Dependent Sentiment Classification*. http://ir.hit.edu.cn/

Wang, Y., Huang, M., Zhao, L., & Zhu, X. (2016). *Attention-based LSTM for Aspect-level Sentiment Classification*.

Werbach, K., & Hunter, D. (2012). *For the Win*.

Xi, N., & Hamari, J. (2019). Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction. *International Journal of Information Management*, *46*, 210–221. https://doi.org/10.1016/j.ijinfomgt.2018.12.002