ANALYZING USER SENTIMENT TO RANK ENGLISH LEARNING APPS ON GOOGLE PLAY

Tran Van Hai[1,*], Bui Khanh Linh[2], Nguyen Duc Long[3]

DOI: http://doi.org/ 10.57001/huih5804.2025.345

ABSTRACT

The rapid growth of EdTech solutions for English language learning has created a saturated market, making it increasingly difficult for users to select appropriate applications. Additionally, this expansion presents challenges for regulators in setting clear standards for product development and circulation within the country. While various frameworks exist to evaluate information systems, few studies have focused on ranking EdTech applications from the user's perspective. Incorporating user feedback into evaluation methods can offer valuable insights for product development, attract investment, and guide users toward better-informed choices. This study proposes a ranking method for English learning apps on the Google Play Store, using an adapted version of the McDelone model, which includes criteria for user satisfaction. User satisfaction was assessed using sentiment analysis on user-generated reviews. The study collected 8,250 reviews from 15 English learning apps available in the Vietnamese market. A deep learning model was used to analyze sentiment and construct a ranking algorithm. For comparison, sentiment prediction was also performed using SVM and Naive Bayes methods. The deep learning approach outperformed traditional models, achieving an RMSE of 0.33 and an R^2 of 0.78, confirming its effectiveness in capturing user sentiment and supporting app ranking efforts.

Keywords: *English Learning Apps, Google Play Store, Sentiment analysis, Deep learning, Edtech.*

[1]Department of Management and Services, Electric Power University, Vietnam
[2]Faculty of Information Technology, Electric Power University, Vietnam
[3]Institute of Information Technology, Vietnam Academy of Science and Technology, Vietnam
[*]Email: haitv@epu.edu.vn
Received: 25/6/2025
Revised: 12/9/2025
Accepted: 28/9/2025

1. INTRODUCTION

In Vietnam, the demand for English language proficiency has been driven by its critical role in education, economic development, and global integration. English has been established as a core subject from primary through secondary education, and the Vietnamese government's National Foreign Language Project 2020 underscores a long-term commitment to improving English competence nationwide [1, 2]. Currently, more than 90% of Vietnamese students learn English as their primary foreign language [2], and enrollment in private English training centers has risen at an annual rate of 30% over the past five years [3, 27]. Correspondingly, the number of English language centers and EdTech solutions, including mobile applications, has expanded rapidly, particularly on digital platforms such as the App Store and Google Play Store [27, 29]. The proliferation of English learning applications on the Google Play Store reflects both the diversity of learner needs and the rapid pace of technological innovation in language education [27, 29]. These applications incorporate advanced technologies such as artificial intelligence (AI), natural language processing (NLP), gamification, and augmented reality (AR) to deliver more personalized, interactive, and effective language learning experiences [31]. As a result, learners benefit from tailored content, real-time feedback, and immersive learning environments that enhance both engagement and outcomes.

Given the growing diversity of English learning applications in Vietnam, choosing an appropriate app has become increasingly difficult for users. Most applications are designed to serve specific age groups and skill levels, making user-centered evaluation essential. However, there is a notable gap in academic research concerning the systematic evaluation of English learning applications. Consequently, many users or parents rely primarily on user-generated reviews to make informed decisions. According to BrightLocal, 87% of consumers trust online reviews as much as personal recommendations, and 93% report that reviews influence their purchasing decisions. Statista further reports that 54% of online users consult reviews to compare product options. However, with the vast volume of reviews available, manually processing this data becomes impractical [30], necessitating the adoption of automated sentiment analysis systems.

Sentiment analysis, a subfield of natural language processing (NLP), applies computational techniques to classify opinions embedded in textual data. Common algorithms include Support Vector Machines (SVM), Decision Trees, Naive Bayes classifiers, and deep learning models [7, 8]. Research by Pang et al. [6] demonstrated that SVM classifiers offer superior performance for sentiment classification tasks, achieving an accuracy of 82.9%. More recently, deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated improved outcomes, with CNNs achieving 90.1% accuracy [4]. BERT is a deep learning model developed by Google that has surpassed prior benchmarks, achieving over 93% accuracy on sentiment analysis tasks [5].

While sentiment analysis in English has matured, its application to Vietnamese data remains limited due to language-specific challenges. Vietnamese text processing faces obstacles such as limited library support and the complexity of word segmentation [32]. Recent studies have attempted to address these issues through customized encoding, preprocessing, and the development of proprietary datasets [33]. However, the multi-step nature of Vietnamese text processing continues to hinder model accuracy, highlighting the need for further research in this area. This study applies a Deep Learning model to enhance the accuracy of sentiment analysis for Vietnamese-language user reviews, utilizing the RapidMiner platform for implementation. Given the limited resources available for Vietnamese, we constructed a custom training dataset and developed a Vietnamese stopword dictionary to support the sentiment classification process. The research focuses on English learning applications for children developed by Vietnamese startups, available on the Google Play Store. A list of English applications was compiled, followed by the collection of 8,250 user reviews using Python-based web scraping. A subset of these reviews was used to train the Deep Learning model, enabling more accurate sentiment detection specific to the Vietnamese language context. First, this study develops bilingual (English and Vietnamese) datasets of reviews. Second, it collects user-generated reviews from 21 English learning apps on the Google Play Store. Third, it proposes a multi-criteria evaluation framework, including System Quality (downloads), User Satisfaction (sentiment analysis), and Information Quality (user ratings). Finally, it implements a Deep Learning-based sentiment analysis model and integrates it with a scoring formula to effectively rank the applications.

The structure of the remainder of this paper is as follows: Section 2 reviews the existing literature on sentiment analysis and information systems evaluation. Section 3 introduces the proposed methodology and evaluation criteria. Section 4 presents the experimental results and model performance. Finally, the conclusion summarizes the findings and offers directions for future research.

2. RELATE WORK

The evaluation of information systems has long been guided by the DeLone and McLean Information System Success Model, first introduced in 1992 and updated in 2003. This model has become one of the most widely adopted frameworks for assessing information system (IS) effectiveness [15, 18, 28]. It outlines six core dimensions: System Quality (technical performance and reliability), Information Quality (accuracy and relevance of output), Service Quality (support provided by the IS team), Use (extent and nature of system usage), User Satisfaction (user perceptions and experiences), and Net Benefits (individual and organizational impact) [10]. Among these, user satisfaction has emerged as a central focus in IS evaluation, as it is strongly correlated with system effectiveness, usage levels, and adoption rates. Numerous studies have confirmed that user satisfaction significantly contributes to assessing the overall success of information systems [9]. This perspective is especially critical in evaluating educational information systems such as E-learning platforms, which have become increasingly prevalent [17]. To adapt the DeLone and McLean model for E-learning environments, researchers have refined and extended the original criteria. For instance, Cidral et al. [12] incorporated additional factors, including instructor attitudes, diversity in assessment methods, and learner interaction. Evaluation often involves statistical modeling techniques such as Partial Least Squares (PLS), with data collected via questionnaires [14, 16]. For example, Efiloğlu [13] conducted a study involving 144 university students in Rome, analyzing survey responses with PLS path modeling. The findings revealed that system quality significantly influenced both system use and user satisfaction, whereas information quality primarily affected satisfaction. Similarly, Marjanovic et al. [15] evaluated Moodle by modifying the DeLone and McLean model, emphasizing system quality, user satisfaction, net benefits, and a newly introduced factor, user performance.

In recent years, the growth of mobile applications has led to a corresponding rise in companies focused on analyzing user behavior and feedback. Analyzing user reviews provides valuable insights into customer emotions, preferences, and expectations, ultimately supporting app developers in enhancing user experiences and promoting wider app adoption [22, 23, 25]. The primary objective of sentiment analysis in this context is to classify user opinions as positive or negative, thereby assisting developers in improving app functionality and guiding companies in strategic decision-making, particularly in monetization strategies related to ratings, reviews, and pricing [20].

Traditional sentiment analysis methods often rely on sentiment lexicons and grammatical rules to detect emotions, but these approaches struggle with handling nuanced expressions or new linguistic trends. Machine learning algorithms such as Naive Bayes, Support Vector Machines (SVM), and Decision Trees have enhanced performance by learning from features like n-grams or bag-of-words. However, these techniques still face challenges in capturing context and semantic depth [21, 24, 26].

The emergence of deep learning has significantly improved sentiment analysis capabilities. Neural networks such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are capable of learning complex language patterns, extracting context-aware representations, and automatically discovering features from large datasets [19, 26]. This study applies deep learning techniques to analyze user sentiments as an indicator of customer satisfaction within a broader evaluation framework.

This study proposes a method to evaluate English learning applications for children developed by Vietnamese startups on the Google Play Store. The evaluation framework is adapted from the DeLone and McLean Information Systems Success Model. Among the criteria, user satisfaction is quantified using sentiment scores derived from user reviews. Detailed methodology and sentiment measurement are presented in Section 3.

3. METHOD

3.1. English learning app evaluation method

This study combines McDelone's information system evaluation with customer sentiment measurement. We propose the following criteria: System Quality (Rating stars), User Satisfaction (Sentiment value), Scalable System (Years of development), and Information Quality (Keyword ranking)

The score of an application is calculated using the formula:

$$\mathrm{Score}(\mathrm{App_i}) = \alpha x_1 + \beta x_2 + \gamma x_3 + \vartheta x_4 \quad (1)$$

Where: $\alpha, \beta, \gamma, \vartheta$ are the coefficients and $\alpha + \beta + \gamma + \vartheta = 1$

x_1, x_2, x_3, x_4 correspond to the five measurement criteria, including: System Quality, User Satisfaction, Scalable System, and Information Quality

The table of criteria and measurement methods is described as shown in Table 1.

Table 1. Criteria for Evaluating English Learning App

Criteria Name	Description	Measurement Method
System Quality	Represented based on the customer rating stars	Number of rating stars
User Satisfaction	Customer satisfaction with the product	Sentiment value
Scalable System	Stability and scalability of the system	Years of development
Information Quality	Number of keywords ranking in the top	Measurement based on tools

3.2. Measuring user satisfaction by Sentiment analysis

In the field of information systems evaluation, assessing an application's performance involves multiple criteria to capture its overall quality and effectiveness. Commonly used criteria include System Quality, Scalable System, and Information Quality. These metrics can typically be gathered through automated analytical tools, offering measurable indicators such as system ratings, development history, and keyword rankings. However, evaluating User Satisfaction presents greater challenges, as it reflects subjective experiences rather than objective metrics. User Satisfaction cannot be directly quantified using conventional methods, as it is often conveyed through user-generated content such as reviews, ratings, and feedback. To gain insight into this emotional and attitudinal data, sentiment analysis becomes essential. Sentiment analysis involves processing natural language to detect positive, negative, or neutral expressions of opinion. In this study, a Deep Learning model is applied to analyze customer sentiment and interpret the emotional tone behind user feedback, as illustrated in Figure 1.

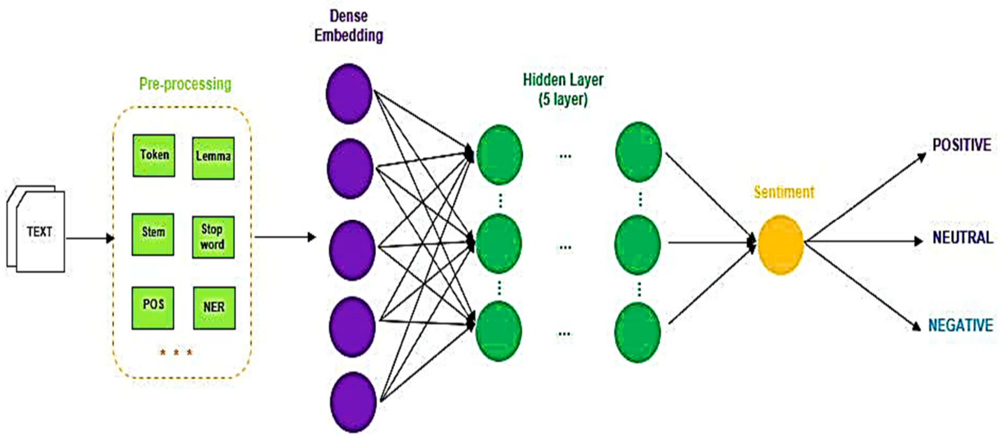

Figure 1. Deep Learning model for sentiment analysis

Sentiment analysis has largely focused on English, while Vietnamese research remains limited due to the lack of suitable tools and resources. To address this gap, we developed a Deep Learning model using the RapidMiner platform, trained on a dataset of nearly 10,000 user reviews from educational apps. Reviews were collected from the Google Play Store, manually labeled as Positive, Negative, or Neutral, and preprocessed using a customized Vietnamese stopword list adapted from English. This helped enhance text cleaning and model performance. We focused on English learning apps developed by Vietnamese companies, collecting reviews and additional app data such as release year, downloads, keyword rankings, and star ratings through analytics tools. The trained model was then used to classify sentiment and calculate sentiment values for each app. Finally, apps were scored based on four key criteria: System Quality, User Satisfaction, Scalable System, and Information Quality, providing an evaluation framework for educational applications in the Vietnamese context.

The sentiment value for each app is calculated using the formula:

$$\text{sentiment score}(\text{App}_i) = \frac{\text{number of positive reviews }(\text{App}_i)}{\text{Total of reviews }(\text{App}_i)} \qquad (2)$$

4. EXPERIMENT

4.1. Data

This study developed three datasets: training data, testing data, and a stopword list. Python was used to collect data from educational apps on the Google Play Store for the training data. After collecting and processing the data and removing icons, we obtained 8,250 cleaned reviews. Next, we collected testing data from English learning apps on the Google Play Store. Apps with fewer than 50 reviews were excluded from the evaluation list. Additionally, for apps with over 2,000 reviews, we filtered out the 550 most recent reviews to measure customer sentiment and ensure data freshness, and the stop word list with 3,324 words. Table 2 lists the 15 selected apps for evaluation using this study method.

Table 2. List English learning apps for evaluation

No	Product's name	Total reviews
1	AI Grammar kiểm tra tiếng Anh	550
2	Bài Test Ngữ Pháp Tiếng Anh	550
3	ELSA Speak_ English Learning	550
4	Eng Breaking_ Game tiếng Anh	550
5	ENGO_ Học giao tiếp tiếng anh	550
6	Fun English Học tiếng Anh	550
7	Học phát âm tiếng Anh TFlat	550
8	Tongo _ Học tiếng Anh	550
9	Học tiếng Anh _ 11.000 từ	550
10	Learn English Grammar	550
11	Lingokids_ Game học tiếng Anh	550
12	Luyen Thi IELTS A _ Z Band 9	550
13	Monkey Junior	550
14	Simpler_ App Học Tiếng Anh	550
15	Tflat	550

4.2. Measuring user satisfaction

- Designing the Deep Learning Model

We applied a Deep Neural Network for sentiment classification, specifically designed for natural language processing tasks. The model includes 5 hidden layers. The model was developed and trained using RapidMiner. Initially, the data was converted from nominal to text format. The document processing stage included text standardization steps such as converting text to lowercase, removing punctuation, and cleaning irrelevant elements. Additionally, a custom stopword list was used to eliminate common, non-informative words (e.g., "and," "but," "is"), making the model focus on meaningful terms. After pre-processing, the data was used to train the deep learning model, optimizing performance and enhancing the quality of the sentiment analysis results.

- Sentiment Score for each English Learning App

Formula (2) was applied to calculate the sentiment value for each English learning app to measure the customer satisfaction criterion. The value of this criterion ranges from [0, 1]. Figure 2 shows the sentiment value for each app.

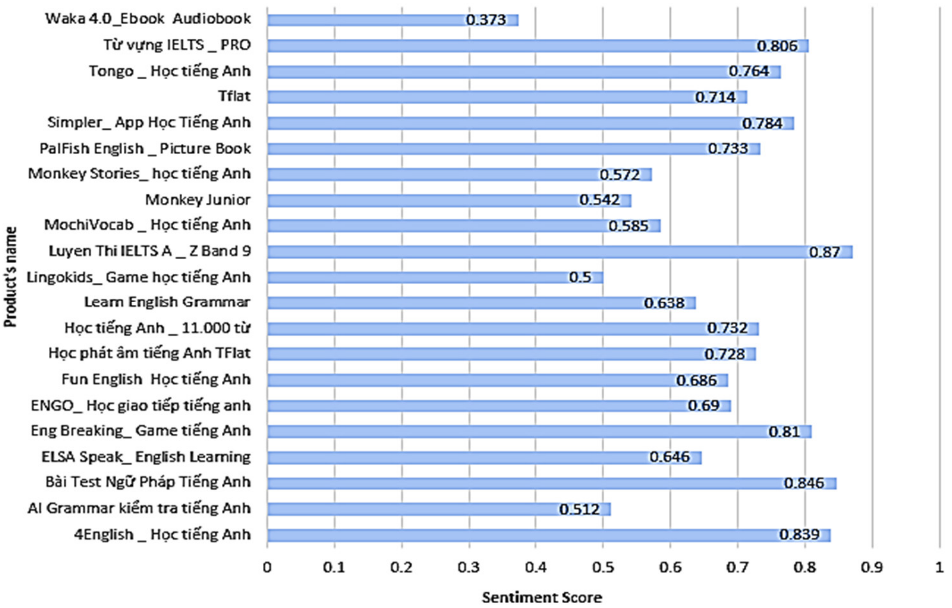

Figure 2. Sentiment score for each English Learning App

Figure 2 shows a significant difference in customer satisfaction levels across various apps. Some apps are highly rated by users, while others receive lower ratings. From the chart, we can see that the apps with the highest customer satisfaction are Luyện Thi IELTS A _ Z Band 9 (0.87), Từ vựng IELTS _ PRO (0.8058), 4English _ Học tiếng Anh (0.8387), Bài Test Ngữ Pháp Tiếng Anh (0.846), and Eng Breaking Game tiếng Anh (0.81).

- Comparison with other machine learning algorithms such as Naive Bayes and SVM

Using the same training set, we compared our model with other algorithms like Naive Bayes and SVM on the same dataset. The average processing time for each algorithm is shown in Table 3.

Table 3. Comparison of the Deep Learning model with Naive Bayes and SVM on the same dataset

Algorithm	**Processing Time**	**Accuracy**
Deep learning	90 -120s	RMSE = 0.33 and R^2 = 0.78
SVM	10 - 20s	Recall = 41.64 and Precision = 45.73
Naive Bayes	1200 - 1300s	Recall = 37.06 and Precision = 58.54

Based on the analysis of processing time and accuracy, it can be concluded that the Deep Learning algorithm outperforms other algorithms when evaluating both criteria simultaneously on the same dataset. Specifically, Deep Learning delivers highly accurate results and maintains efficient processing time, demonstrating its superiority in handling complex datasets. In contrast, traditional machine learning algorithms often struggle to optimize both criteria at the same time.

4.3. Ranking results

After measuring the values for the four criteria, we used formula (1) to calculate the scores for the apps. The final scores of the apps are shown in Figure 3.

After evaluating the scores, the Lingokids Games English learning app ranked the highest. This indicates that the game-based approach to learning English is the most effective and easy to absorb. Moreover, this English product is aimed at children, showing that the development of English apps for the children's segment receives the most attention. Following closely is Elsa Speak, ranked second in the list based on scores. This product is highly rated in practice, and when applying the formula, it also ranked well. This demonstrates that the measurement approach based on our proposed criteria is appropriate and provides good accuracy, including the customer satisfaction measurement algorithm.

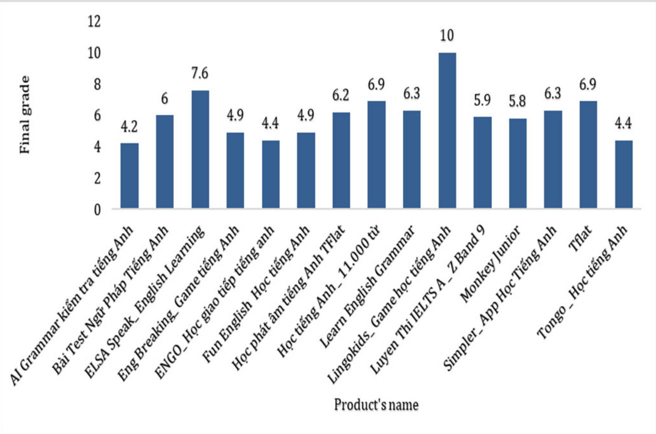

Figure 3. Final Score of English Learning Apps

5. CONCLUSION

Many existing studies evaluating information systems are based on the DeLone and McLean (McDelone) model. These studies typically apply both qualitative and quantitative methods, with surveys being the most common approach to measure user satisfaction. In this study, we enhance the traditional DeLone model by proposing a set of four evaluation criteria: System Quality, Scalable System, Information Quality, and User Satisfaction. To ensure objectivity and efficiency, we apply quantitative methods using application analysis tools to measure System Quality, Scalable System, and Information Quality. For User Satisfaction, we employ sentiment analysis techniques based on user reviews. This model was trained and tested using a dataset of user reviews and compared with traditional machine learning methods such as Support Vector Machines (SVM) and Naive Bayes. The comparison shows that the deep learning model significantly outperforms traditional methods in both accuracy and processing time, demonstrating its effectiveness in sentiment classification for application evaluation. This integrated approach offers a more automated, scalable, and reliable method for assessing user experience and system performance.

REFERENCES

[1]. Nguyen T., "The Role of English in Vietnamese Education," *Journal of Education and Practice*, 2020.

[2]. Pham H., Do T., "Economic Impacts of English Proficiency in Vietnam," *Asian Economic Review*, 2019.

[3]. Le A., "English as a Tool for Global Integration in Vietnam," *Cultural Studies Journal*, 2021.

[4]. Zhang L., Wang S., Liu B., "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), 2018.

[5]. Devlin J., Chang M. W., Lee K., Toutanova K., "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019.

[6]. Pang B., Lee L., Vaithyanathan S., "Thumbs up? Sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, 2002.

[7]. Go A., Bhayani R., Huang L., "Twitter sentiment classification using distant supervision," *CS224N Project Report, Stanford*, 2009.

[8]. Pak A., Paroubek P., "Twitter as a corpus for sentiment analysis and opinion mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, 2010.

[9]. Al-Fraihat D., Joy M., Sinclair J., "Evaluating E-learning systems success: An empirical study," *Computers in Human Behaviour*, 102, 67-86, 2020.

[10]. DeLone W. H., McLean E. R., "Information systems success measurement," *Foundations and Trends® in Information Systems*, *2*(1), 1-116, 2016.

[11]. Dwivedi Y. K., et al., "Research on information systems failures and successes: Status update and future directions," *Information Systems Frontiers*, 17, 143-157, 2015.

[12]. Cidral W. A., Oliveira T., Di Felice M., Aparicio M., "E-learning success determinants: Brazilian empirical study," *Computers & education*, 122, 273-290, 2018.

[13]. Efiloğlu Kurt Ö., "Examining an e-learning system through the lens of the information systems success model: Empirical evidence from Italy," *Education and Information Technologies*, 24(2), 1173-1184, 2019.

[14]. Alyoussef I. Y., "Acceptance of e-learning in higher education: The role of task-technology fit with the information systems success model," *Heliyon*, 9(3), 2023.

[15]. Marjanovic U., Delić M., Lalic B., "Developing a model to assess the success of e-learning systems: evidence from a manufacturing company in transitional economy," *Information Systems and e-Business Management*, 14, 253-272, 2016.

[16]. Çelik K., Ayaz A., "Validation of the Delone and McLean information systems success model: a study on student information system," *Education and Information Technologies*, 27(4), 4709-4727, 2022.

[17]. Galvis Á. H., Carvajal D., "Learning from success stories when using eLearning and bLearning modalities in higher education: a meta-analysis and lessons towards digital educational transformation," *International Journal of Educational Technology in Higher Education*, 19(1), 23, 2022.

[18]. Jeyaraj A., "DeLone & McLean models of information system success: Critical meta-review and research directions," *International Journal of Information Management*, 54, 102139, 2020.

[19]. Venkatakrishnan S., Kaushik A., Verma J. K., "Sentiment analysis on google play store data using deep learning," *Applications of Machine Learning*, 15-30, 2020.

[20]. Li Y., Jia B., Guo Y., Chen X., "Mining user reviews for mobile app comparisons," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3), 1-15, 2017.

[21]. Kumar A., Chakraborty S., Bala P. K., "Text mining approach to explore determinants of grocery mobile app satisfaction using online customer reviews," *Journal of Retailing and Consumer Services*, 73, 103363, 2023.

[22]. Liang T. P., Li X., Yang C. T., Wang M., "What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach," *International Journal of Electronic Commerce*, 20(2), 236-260, 2015.

[23]. Tian H., Grover V., Zhao J., Jiang Y., "The differential impact of types of app innovation on customer evaluation," *Information & Management*, 57(7), 103358, 2020.

[24]. Taju S. W., Putra E. Y., Mandias G. F., "Sentiment Identification System for E-Commerce Mobile App Reviews Using Single Layer Neural Network," In *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)* (pp. 1-7), IEEE, 2022.

[25]. Liu H., Wu S., Zhong C., Liu Y., "The effects of customer online reviews on sales performance: The role of mobile phone's quality characteristics," *Electronic Commerce Research and Applications*, 57, 101229, 2023.

[26]. Ullah W., Zhang Z., Stefanidis K., "Sentiment analysis of mobile apps using BERT," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 66-78). Cham: Springer Nature Switzerland, 2023.

[27]. Nguyen H., Quyet P. D., "Ranking English Apps From the Google Play Store Based on Customer Reviews Analysis," *International Journal of Asian Business and Information Management (IJABIM)*, 15(1), 1-18, 2024.

[28]. Thu H. N. T., Khanh L. B., Xuan T. N., "Ranking E-learning Systems in Vietnamese K12 Market Based on Multiple Criteria," in *Intelligent Systems and Networks: Selected Articles from ICISN 2023, Vietnam*, 752, 435, 2023.

[29]. Tommerdahl J. M., Dragonflame C. S., Olsen A. A., "A systematic review examining the efficacy of commercially available foreign language learning mobile apps," *Computer Assisted Language Learning*, 37(3), 333-362, 2024.

[30]. Griffith S. F., Hagan M. B., Heymann P., Heflin B. H., Bagner D. M., "Apps as learning tools: a systematic review," *Pediatrics*, 145(1), 2020.

[31]. Berns A., "A review of virtual reality-bassed language learning apps," *RIED. Revista iberoamericana de educación a distancia*, 2021.

[32]. Vo Q. T., Tran P., Tran T., "Sentiment analysis for a low-resource language: A study on a Vietnamese University," *ICIC Exp. Lett*, *14*, 1115-1124, 2023.

[33]. Quoc Tran K., Trong Nguyen A., Hoang P. G., Luu C. D., Do T. H., Van Nguyen K., "Vietnamese hate and offensive detection using PhoBERT-CNN and social media streaming data," *Neural Computing and Applications*, 35(1), 573-594, 2023.