JSLHU
JOURNAL OF SCIENCE
OF LAC HONG UNIVERSITY

ISSN: 2525 - 2186
Tạp chí Khoa học Lạc Hồng, **2026**, 26, 037-042

HIỂU VỀ VIỆC CHUYỂN ĐỔI NỀN TẢNG MẠNG XÃ HỘI THÔNG QUA CÁC YẾU TỐ CẢM XÚC VÀ THÓI QUEN SỬ DỤNG

Trần Hữu Duật *

Trường Đại học Thủ Dầu Một, Số 06 Trần Văn Ơn, Phường Phú Lợi, Hồ Chí Minh, Việt Nam

* Tác giả liên hệ: *duatth@tdmu.edu.vn*

THÔNG TIN BÀI BÁO

Ngày nhận:	06/03/2026
Ngày hoàn thiện:	27/03/2026
Ngày chấp nhận:	04/04/2026
Ngày đăng:	15/04/2026

TỪ KHÓA

Sentiment analysis;
Support Vector Machine;
Social media platform;
Threads;
Usage factor.

TÓM TẮT

Nghiên cứu này phân tích cảm xúc của người dùng và các yếu tố sử dụng ứng dụng trong bối cảnh cạnh tranh giữa X (trước đây là Twitter) và Threads của Meta. Các đánh giá của người dùng từ Google Play Store và Apple App Store được thu thập và phân tích bằng phương pháp Support Vector Machine (SVM) để phân loại cảm xúc. Kết quả phân loại sau đó được phân tích tương quan với các yếu tố sử dụng ứng dụng chính, bao gồm khả năng sử dụng (usability), tính năng (features), thiết kế (design), và hỗ trợ (support). Kết quả nghiên cứu cho thấy cảm xúc của người dùng có ảnh hưởng đáng kể đến mức độ tương tác với nền tảng. Sự không hài lòng ban đầu đối với X đã góp phần tạo ra cảm xúc tích cực đối với Threads; tuy nhiên, khi mức độ tương tác với Threads giảm sút, người dùng dần quay trở lại sử dụng X. Phân tích tương quan cũng cho thấy mối quan hệ tương quan âm giữa các yếu tố tương tự trên hai nền tảng, qua đó phản ánh động lực cạnh tranh giữa các ứng dụng mạng xã hội.

UNDERSTANDING SOCIAL MEDIA PLATFORM SWITCHING THROUGH SENTIMENT AND USAGE FACTORS

Tran Huu Duat *

Thu Dau Mot University, No. 6, Tran Van On Street, Phu Loi Ward, Ho Chi Minh City, Vietnam

* Corresponding author: *duatth@tdmu.edu.vn*

ARTICLE INFO

Received:	Mar 6th, 2026
Revised:	Mar 27th, 2026
Accepted:	Apr 4th, 2026
Published:	Apr 15th, 2026

KEYWORDS

Sentiment analysis;
Support Vector Machine;
Social media platform;
Threads;
Usage factor.

ABSTRACT

This study analyzes user sentiment and application usage factors in the competition between X (formerly Twitter) and Meta's Threads. User reviews from the Google Play Store and Apple App Store were examined using Support Vector Machine (SVM) for sentiment classification. The results were correlated with key usage factors: usability, features, design, and support. Findings show that user sentiment significantly influences engagement. Initial dissatisfaction with X contributed to positive sentiment toward Threads. However, as Threads' engagement declined, users gradually returned to X. Correlation results also indicate negative relationships between similar factors across the platforms, highlighting competitive dynamics between social media applications.

Doi: https://doi.org/10.61591/jslhu.26.1088
Available online at: https://lhj.vn

1. INTRODUCTION

The digital era has seen rapid growth of social media platforms competing for global user attention. Threads, launched by Meta as a direct competitor to X (formerly Twitter), aimed to leverage instability within X and build a large-scale platform for global conversations [1]. Shortly after its launch, Threads gained over five million downloads within four hours, becoming the fastest-growing social media application, partly driven by user dissatisfaction with policy changes on X. However, this momentum declined quickly, with daily active users dropping significantly and average usage falling to about three minutes per day. Meanwhile, X remained stable with around 100 million daily users and approximately twenty-five minutes of daily engagement [2]. These contrasting trends highlight the need to understand factors influencing user retention and engagement across competing social media platforms [3]. Understanding these dynamics requires analyzing user perceptions expressed through online reviews on platforms such as the Google Play Store and Apple App Store. These reviews provide insights into user satisfaction, usability, and reasons for platform switching. However, limited research has examined how shifts in user sentiment influence competition between new and established social media platforms, or which specific application features shape these sentiments.

Sentiment analysis, often referred to as opinion mining, represents a research field devoted to examining individuals' attitudes, judgments, and emotional responses toward various entities such as products, services, organizations, or events. The discipline encompasses several related concepts, including opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining, which share the objective of interpreting subjective information embedded within textual data [4]. In the context of modern digital communication, sentiment analysis has become a crucial tool for understanding public opinion within online environments, particularly across social networks where users continuously generate large volumes of evaluative content [5].

Advancements in machine learning techniques have significantly enhanced the capacity to analyze such textual information. Among these approaches, Support Vector Machine (SVM) has become one of the most widely adopted algorithms for sentiment classification due to its strong performance in handling high-dimensional text data. To further improve model effectiveness, feature selection strategies are frequently applied to identify the most relevant attributes within textual datasets. One commonly used method is the chi-square technique, which evaluates the statistical relationship between terms and sentiment categories in order to select informative features and reduce computational complexity [6].

The analytical process underlying sentiment analysis is closely connected to text mining, which involves extracting meaningful patterns and knowledge from large collections of textual data. Text mining is widely applied in areas such as document clustering, text classification, and sentiment evaluation. Typically, the process consists of several sequential stages, including dataset collection, text preprocessing, feature representation, feature selection, classification, and evaluation [7]. The authors in [8] describe a typical classification-based sentiment analysis workflow beginning with dataset acquisition, followed by preprocessing procedures such as tokenization, stop-word removal, and stemming. The text data are then transformed into numerical representations through weighting techniques before undergoing feature selection to eliminate irrelevant attributes. Classification algorithms, such as naive Bayes, K-nearest neighbor (KNN), or SVM are subsequently applied, and the model's performance is evaluated using metrics such as accuracy and the area under the curve (AUC).

Among these algorithms, SVM has consistently demonstrated strong predictive capability in both classification and regression tasks. Unlike probabilistic classifiers, SVM focuses on maximizing the margin between categories by determining a hyperplane that best distinguishes between labeled data instances [9]. The algorithm seeks a separating surface capable of distinguishing these classes effectively within the feature space [10].

Beyond classification, statistical techniques such as correlation analysis are often used to explore relationships between variables extracted from textual data. Correlation refers to the statistical association between two variables, most commonly describing the degree to which they exhibit a linear relationship [11]. One widely applied measure is the Pearson correlation coefficient, which quantifies the strength and direction of the linear relationship between two variables. Even when relationships involve certain nonlinear characteristics, Pearson correlation can still reveal meaningful associations within the data. In exploratory research, correlation analysis serves as an important tool for interpreting relationships and comparing results across studies.

Several previous studies have compared machine learning algorithms to determine the most effective method for sentiment classification tasks. The author et al. in [12] analyzed public sentiment data and compared naive Bayes, SVM, and decision tree algorithms. Their results showed that SVM achieved outperforming naive Bayes, and decision tree. Similarly, in [13] evaluated sentiment classification using the sentiment dataset and found that SVM achieved an accuracy exceeding the performance of KNN and naive Bayes.

Across previous studies, SVM consistently demonstrates strong performance in sentiment classification across various contexts, including policy discussions and product evaluations. Due to its reliability and predictive capability, this study adopts SVM as the primary model for analyzing user sentiment toward Threads and X. User reviews from the Google Play Store and Apple App Store are classified into positive, negative, and neutral sentiments. The results are then associated with key application usage factors: usability, features, design, and support to identify elements influencing user perception and engagement. By integrating sentiment analysis with application usage factors, this research aims to provide insights into how user perceptions shape engagement patterns in competing social media platforms, offering practical implications for developers to improve features and user experience.

2. Materials and Methods

This study adopts a structured methodology to collect, process, and analyze user reviews related to the Threads and X applications obtained from the Google Play Store and Apple App Store. The methodological framework is adapted from Amrie et al. [19], with modifications to suit the context of analyzing competing social media platforms. The overall stages of the research procedure are conceptually illustrated in Figure 1.

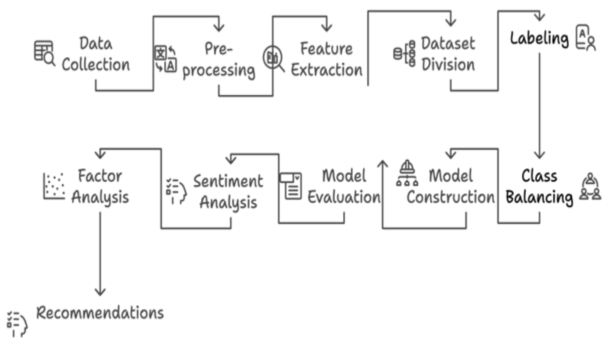

Figure 1: The workflow research

The research begins with collecting user reviews from the two mobile application marketplaces using web scraping techniques. Two Python-based tools, google-play-scraper and app-store-scraper, are utilized to retrieve review texts and their associated metadata. The collected reviews are grouped according to the application and platform source to enable comparative analysis between Threads and X as well as between Google Play and the App Store. All reviews included in the dataset are written in English and are consolidated into a structured dataset using the pandas library.

To label sentiment automatically, this study applies a rating-based labeling approach. Reviews with ratings greater than three (>3) are classified as positive, ratings equal to three (=3) are labeled neutral, and ratings below three (<3) are categorized as negative. This labeling process is implemented using Python scripts with the pandas library to ensure consistency across the dataset.

Before classification, the review texts undergo several preprocessing steps to improve data quality. These steps are implemented using the Natural Language Toolkit (NLTK) and the demoji library. The preprocessing process includes text cleaning to remove irrelevant symbols and emojis, case folding to convert all characters to lowercase, tokenization to split sentences into individual words, stop-word removal to eliminate commonly used words with little analytical value, and stemming to reduce words to their root forms.

Tokenization divides each review into tokens to facilitate text analysis and improve the efficiency of NLP algorithms [20]. During this stage, punctuation is removed and all characters are standardized to lowercase. Stop-word removal eliminates frequently occurring but semantically weak words such as is, an, and the, thereby reducing noise and improving computational efficiency [20]. Stemming is then applied to reduce words to their base forms by removing prefixes or suffixes, which helps reduce vocabulary size and improve processing efficiency [9].

After preprocessing, feature extraction is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method implemented through the scikit-learn library. TF-IDF assigns weights to words based on their importance within a document relative to the entire dataset, allowing the model to emphasize more informative terms for sentiment classification [21].

Finally, sentiment classification is conducted using the SVM algorithm. SVM is a supervised learning method that constructs an optimal decision boundary known as the maximum marginal hyperplane (MMH) to separate data points into different classes. To address non-linearly separable data, SVM employs the kernel trick, which maps data into a higher-dimensional space where separation becomes possible. Due to its ability to handle high-dimensional textual data effectively, SVM is widely used in sentiment analysis tasks and provides strong classification performance [21].

In the result, model evaluation is conducted to assess the performance of the implemented classification model. The evaluation uses accuracy, precision, recall, and F1-score metrics. Accuracy indicates the proportion of correctly classified instances, while precision measures the correctness of positive predictions for a specific class.

Recall evaluates the model's ability to identify all relevant instances of a class. The F1-score provides a balanced measure by combining precision and recall. In this study, the performance of the SVM model is evaluated using the scikit-learn library.

Furthermore, after completing the sentiment analysis on the testing data of X and Threads application reviews, the next step is to analyse the relationship between the sentiment results and application usage factors. In this study, the usage factors are categorized into four main aspects: usability, features, design, and support. Usability refers to the ease of use and efficiency of an application and plays a crucial role in shaping early user experiences [22]. Application features can enhance user engagement and encourage repeated usage [23]. Meanwhile, attractive and intuitive design contributes significantly to improving user satisfaction [24]. In addition, effective customer support helps resolve user issues efficiently and strengthens user trust and loyalty [25]. These four factors collectively influence the overall user experience.

Table 1 presents the keyword lists representing each usage factor category, constructed using the NLTK corpus lemmas and synsets library. Each review is examined to determine whether it contains keywords related to usability, features, design, or support. The identification process utilizes lexical relationships such as hypernyms, hyponyms, meronyms, and holonyms from the NLTK corpus. If a match is identified, the review is associated with the corresponding factor and labelled according to its detected sentiment (positive, negative, or neutral).

To analyse the relationship between sentiment and application usage factors, correlation analysis is performed using the Pearson correlation method implemented in the pandas library. The correlation results are visualized using the matplotlib library. All analytical processes are implemented through Python scripts to ensure accuracy, reproducibility, and clarity in interpreting the results.

Table 1. *List of keywords for application usage factors*

Factors	Keywords	Subcategories
Usability	function, application, utility, enjoy, useful, usability, usable, easy, simple, efficient, performance, accessibility	positive, neutral, negative
Features	features, feed, product, profile, trending, update, tools, options, functionality, content, notification	positive, neutral, negative
Design	design, interface, interaction, navigation, experience, layout, appearance, visual, color, theme, display	positive, neutral, negative
Support	support, help, service, assistance, response, feedback, problem, issue, customer service, complaint, resolution	positive, neutral, negative

3. RESULTS AND DISCUSSION

This section presents the analysis of the research findings based on the methodology described previously. The primary objective is to identify the factors contributing to the rapid decline in the usage of the Threads social media application. In addition, the study aims to understand the reasons why users tend to return to using Twitter (X) after initially adopting Threads.

3.1 Data processing

User reviews for the Threads and X applications were collected through web scraping from the Google Play Store and Apple App Store. After the collection process, the reviews were grouped according to the respective applications and combined into a unified dataset for each platform. Several data engineering procedures were then applied, including data cleaning to remove noise and irrelevant information, followed by text mining processes such as tokenization, stop-word removal, emoji removal, stemming, and text weighting. These preprocessing steps were implemented using the NLTK and demoji libraries to prepare the textual data for sentiment analysis.

The dataset consists of 179,622 raw review entries for the X application and 39,316 raw review entries for the Threads application, which form the basis for subsequent sentiment classification and factor correlation analysis.

The training dataset derived from user reviews of the X and Threads applications was labeled based on the rating scores provided by users on the Google Play Store and Apple App Store. Reviews were categorized into positive (>3), neutral (=3), and negative (<3) sentiment classes. The distribution of these sentiment labels is presented in Figure 2, illustrating the overall sentiment trends for both applications.

The results indicate that X contains a higher proportion of negative reviews compared to positive and neutral ones, suggesting a level of user dissatisfaction. In contrast, Threads demonstrates a relatively more balanced sentiment distribution with a considerable number of positive reviews, indicating a more favourable initial reception among users. This sentiment distribution serves as an important basis for further analysis of user behaviour and engagement patterns, particularly in understanding the factors influencing the decline in Threads usage and the tendency of users to return to X after previously adopting Threads.

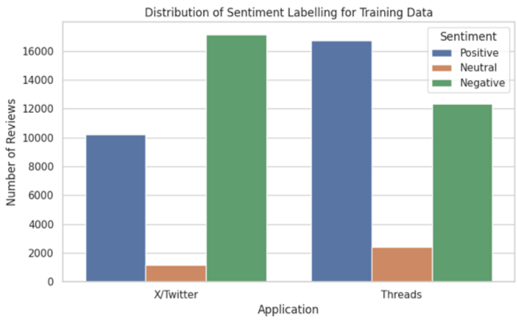

Figure 2: Distribution of Sentiment Labelling for Data

3.2 Sentiment Analysis

Sentiment analysis in this study utilizes the Support Vector Machine (SVM) classification method to categorize user reviews into positive, neutral, and negative sentiments. The model performance is evaluated using accuracy, precision, recall, and F1-score metrics, with the results summarized in Figure 3. The results indicate that the SVM model performs well in identifying positive and negative sentiments for X, achieving F1-scores of 81% and 88%, respectively. These results suggest that the model is effective in distinguishing clear expressions of satisfaction and dissatisfaction. However, the model fails to detect neutral sentiments, as reflected by the 0% precision, recall, and F1-score for this class.

This limitation is likely caused by the small number of neutral reviews and the ambiguous linguistic patterns typically present in neutral statements. For Threads, the model demonstrates a high recall of 93% for positive sentiments, indicating that most positive reviews are correctly identified. However, the precision value of 76% shows that some non-positive reviews are incorrectly classified as positive.

Similar to the results for X, the model struggles to recognize neutral sentiments in Threads reviews, with a recall of only 3%. This consistent difficulty suggests that neutral sentiment classification remains a challenge for the model across both platforms.

Although negative sentiment detection in Threads is relatively effective, its performance is slightly lower compared to X. This difference may be related to less distinctive linguistic cues used by Threads users when expressing dissatisfaction. Overall, these findings highlight the presence of class imbalance in the dataset and indicate the need for improved data balancing or enhanced feature extraction techniques to improve classification performance, particularly for neutral sentiments.

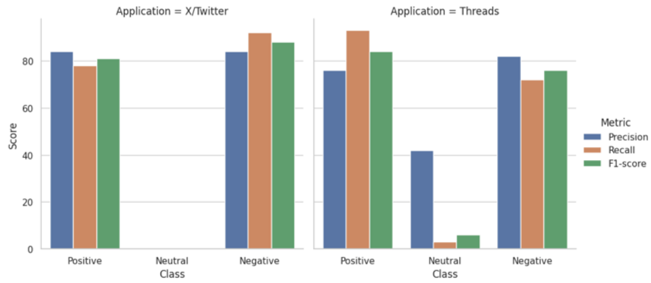

Figure 3: Classification Performance

Furthermore, the testing results presented in Figure 4 reveal that reviews for X are dominated by negative sentiments, which aligns with the patterns observed in the training data. In contrast, Threads reviews show a higher proportion of positive sentiments, although neutral and negative opinions are still present. These results reflect differences in user experiences between the two platforms and suggest potential areas for improvement in both application development and sentiment detection methods.

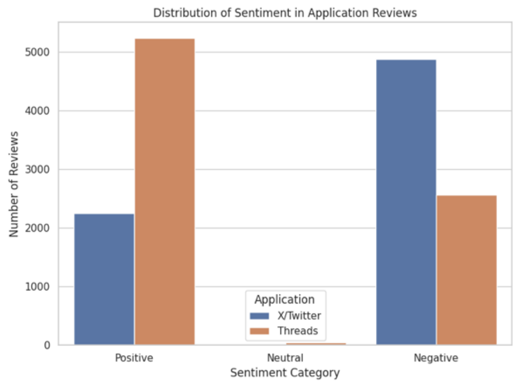

Figure 4: Sentiment result in the application reviews

4. CONCLUSION

This study highlights the competitive interaction between X and Threads based on user sentiment analysis. Initial dissatisfaction with X encouraged users to adopt Threads, but declining engagement with Threads led to the recovery of X's user base. The findings show that improvements in usability, features, design, and support on one platform may influence negative perceptions of the competing platform. Despite providing useful insights, the study is limited by dataset coverage and the challenges of capturing complex user emotions. Future research should incorporate larger datasets and more advanced analytical methods to better understand user engagement and platform sustainability.

5. REFERENCES

[1] Zhang, X., et al., (2025). Natural language processing and text mining in transportation: Current status, challenges, and future roadmap. Expert Systems with Applications, 129050.

[2] Luttrell, R. (2025). Social media: How to engage, share, and connect. Bloomsbury Publishing USA.

[3] Ahmad, H., Khan, N., & Shahid, S. (2025). Mitigating Toxicity in Social Media: Redesign Guidelines for Cultivating Positive User Interactions in the Instagram Threads App. In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, pp. 1-7.

[4] Parveen, N., et al., (2023). Twitter sentiment analysis using hybrid gated attention recurrent network. Journal of Big Data, 10(1), 50.

[5] Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A survey of sentiment analysis from social media data. IEEE Transactions on Computational Social Systems, 7(2), 450-464.

[6] Logan, A. P., LaCasse, P. M., & Lunday, B. J. (2023). Social network analysis of Twitter interactions: a directed multilayer network approach. Social Network Analysis and Mining, 13(1), 65.

[7] Shaik, T., et al., (2023). Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal, 2, 100003.

[8] Dhammananda, J., Budi, I., & Santoso, A. B. (2023). Sentiment Analysis and Topic Modeling of E-Grocery Application Reviews Using Naive Bayes and Support Vector Machine: A Case Study of Segari Data Review on the Google Play Store. In 2023 3rd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), pp. 13-18.

[9] Sanavio, C., Tignone, E., & Ercolessi, E. (2023). Entanglement classification via witness operators generated by support vector machine. The European Physical Journal Plus, 138(10), 936.

[10] Aldisa, R. T., & Maulana, P. (2022). Analisis Sentimen Opini Masyarakat Terhadap Vaksinasi Booster COVID-19 Dengan Perbandingan Metode Naive Bayes, Decision Tree dan SVM. Building of Informatics, Technology and Science (BITS), 4(1), 106-109.

[11] Akus, O. K., Chika, A. F., Ekene, B., & Inono, O. C. (2025). Correlation Assessment of Air Quality Parameters and Climate Variables in Benin City, Southern Nigeria: A Remote Sensing Approach. Dutse Journal of Pure and Applied Sciences, 11(3d), 257-272.

[12] Ayash, L., Algarni, A., & Alqahtani, O. (2025). Advancements in feature selection and extraction methods for text mining: a review. Discover Applied Sciences, 7(8), 914.

[13] Dey, R., Ray, A. K., Srivastava, A., Sharma, D. K., & Piri, J. (2025, February). Sentiment analysis using machine learning techniques with a digital dataset survey. In 2025 International Conference on Emerging Systems and Intelligent Computing (ESIC), pp. 410-414.

[14] Amrie, S., Kurniawan, S., Windiatmaja, J. H., & Ruldeviyani, Y. (2022, February). Analysis of Google Play Store's Sentiment Review on Indonesia's P2P Fintech Platform. In 2022 IEEE Delhi Section Conference (DELCON), pp. 1-5.

[15] Kumar, S., Jailani, N. A., Singh, A. R., & Panchal, S. (2022, April). Sentiment analysis on online reviews using machine learning and NLTK. In 2022 6th international conference on trends in electronics and informatics (ICOEI), pp. 1183-1189.

[16] Rodrigues, G. N., Mir, M. N. H., Bhuiyan, M. S. M., Rafi, M. D. A. L., Hoque, A. M., Maua, J., & Mridha, M. F. (2025). NLP-driven customer segmentation: A comprehensive review of methods and applications in personalized marketing. Data Science and Management

[17] Alsanousi, B., Albesher, A. S., Do, H., & Ludi, S. (2023). Investigating the user experience and evaluating usability issues in ai-enabled learning mobile apps: An analysis of user reviews. International Journal of Advanced Computer Science and Applications, 14(6).

[18] Mehnaz, Jin, J., & Bhasin, A. (2021). Impact of social media technologies on consumers' online buying tendencies with moderating role of gender. International Journal of Applied Decision Sciences, 14(6), 731-744.

[19] Won, D., Chiu, W., & Byun, H. (2023). Factors influencing consumer use of a sport-branded app: the technology acceptance model integrating app quality and perceived enjoyment. Asia Pacific Journal of Marketing and Logistics, 35(5), 1112-1133.

[20] Adeniran, I. A., Efunniyi, C. P., Osundare, O. S., & Abhulimen, A. O. (2024). Transforming marketing strategies with data analytics: A study on customer behavior and personalization. International Journal of Management & Entrepreneurship Research, 6(8), 41-51.