# SOCIAL NETWORK AND SENTIMENT ANALYSIS FOR ENHANCING SOCIAL CRM IN INDONESIAN EDUCATIONAL TECHNOLOGY PLATFORMS

**Rifaa Khairunnisa-[1], Johannes Hamonangan Siregar-[2]***

Sistem Informasi, Fakultas Teknologi dan Desain
Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia
rifaa.khairunnisa@student.upj.ac.id-[1] , johannes.siregar@upj.ac.id-[2]*

*Corresponding Author

**Abstract**

The rapid advancement of digital technology has significantly transformed the education sector, including in Indonesia. According to the 2024 report by Badan Pusat Statistik (BPS), e-learning is among the primary reasons Indonesians access the internet. This trend has positioned educational technology (EdTech) platforms such as Ruangguru, Pahamify, and Zenius as key players in the country's e-learning ecosystem. Simultaneously, social media has become a space where users actively express their experiences regarding the services they use. This study aims to examine user interaction dynamics and public sentiment toward these three EdTech platforms through an integrated approach combining Social Network Analysis (SNA) and Lexicon-Based Sentiment Analysis. Data were collected from platform X and preprocessed for analysis. Network analysis used Gephi to evaluate structural properties and centrality measures, while sentiment analysis used a combination of the InSet lexicon and user-generated vocabulary. To further capture discussion themes, topic modeling using the BERTopic algorithm was also applied to categorize dominant topics from user conversations. The results show that each platform exhibits different social network characteristics. Zenius demonstrates efficient information flow, Ruangguru displays tightly connected user interactions, and Pahamify presents a more dispersed structure. Overall, the sentiment analysis showed that Ruangguru and Zenius had relatively higher proportions of positive sentiment, with 44.6% and 41.4%, respectively. These findings highlight how integrating SNA and sentiment analysis can form a strong foundation for developing Social CRM strategies to enhance the quality of digital education services in Indonesia.

Keywords: Social Network Analysis; Sentiment Analysis; Lexicon-Based; Educational Technology; Social CRM

***Abstrak***

*Kemajuan teknologi digital mendorong transformasi besar dalam sektor pendidikan, termasuk di Indonesia. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, pembelajaran daring merupakan salah satu alasan utama masyarakat mengakses internet. Hal tersebut memperkuat posisi platform teknologi pendidikan (EdTech) seperti Ruangguru, Pahamify, dan Zenius sebagai aktor penting dalam ekosistem pendidikan digital nasional. Di sisi lain, media sosial telah menjadi ruang ekspresi bagi pengguna untuk menyampaikan pengalaman terhadap layanan yang mereka gunakan. Penelitian ini bertujuan untuk memahami dinamika interaksi dan opini pengguna terhadap tiga platform EdTech melalui integrasi metode Social Network Analysis (SNA) dan analisis sentimen menggunakan metode Lexicon-Based. Data dikumpulkan dari platform X dan diproses untuk keperluan analisis. Analisis jaringan dilakukan menggunakan Gephi untuk mengukur properti jaringan dan centrality, sedangkan analisis sentimen memanfaatkan kombinasi kamus InSet dan kosakata UGC. Untuk menangkap tema diskusi yang lebih spesifik, pemodelan topik dengan algoritma BERTopic juga digunakan guna mengelompokkan topik dominan dari percakapan pengguna. Hasil menunjukkan bahwa tiap platform memiliki karakteristik jaringan sosial yang berbeda. Zenius memiliki jaringan efisien dalam sebaran informasi, Ruangguru memperlihatkan koneksi yang padat antar pengguna, dan Pahamify cenderung memiliki struktur yang tersebar. Sentimen pengguna secara umum menunjukkan kecenderungan positif pada Ruangguru dan Zenius, yaitu sebesar 44,6% dan 41,4%. Temuan tersebut menunjukkan bahwa integrasi SNA dan analisis sentimen dapat menjadi dasar kuat dalam penyusunan strategi Social CRM yang relevan untuk meningkatkan kualitas layanan pendidikan di Indonesia.*

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.383
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

*Kata kunci: Analisis Jejaring Sosial; Analisis Sentimen; Lexicon-Based; Educational Technology; Social CRM*

## INTRODUCTION

The rapid development of digital infrastructure has transformed the way people interact, learn, and access services. In Indonesia, the advancement of digital connectivity has led to a significant shift in how educational services are delivered and consumed. This transformation was accelerated by the COVID-19 pandemic, which forced educational institutions and learners to adopt remote and technology-based learning models. The transformation of educational services has led EdTech platforms to rely more heavily on user data and analytics to remain agile and sustainable, especially in post-pandemic learning ecosystems (Komljenovic, Birch, & Sellar, 2024).

One of the critical outcomes of this shift is the emergence of Educational Technology (EdTech) platforms such as Ruangguru, Pahamify, and Zenius. These platforms serve as key actors in bridging the gap between educational needs and technological innovation. As digital education becomes more prevalent, these platforms are not only evaluated based on their academic content but also on how effectively they engage and respond to their users, a factor shaped by effective social media strategies that strengthen emotional connection and increase the intention to adopt EdTech platforms (Kadek, Divananda, & Rubiyanti, 2024).

According to the 2024 report from Badan Pusat Statistik (BPS), online learning ranks as the fourth most common reason Indonesians use the internet, after social media, communication, and entertainment. This trend reflects an increasing dependence on digital learning and affirms the need for continuous evaluation of EdTech services from a user-centered perspective. Evaluating user perception is crucial, as it directly influences learning effectiveness and user satisfaction in digital learning platforms (Sabilla & Hartarto, 2024).

EdTech platforms operate within a user-centric ecosystem, where feedback, reviews, and public opinion are no longer confined to formal surveys or interviews but are increasingly captured organically on social media. These digital footprints, commonly referred to as User-Generated Content (UGC), include tweets, mentions, and replies. However, the unstructured nature and high volume of UGC present analytical challenges that require advanced methodologies. Text-based social media content presents particularly unique challenges for established data analytical approaches (Chadwick, Parry, Ahmed, & Fenton, 2023). This is due to the informal language, inconsistent structure, and large volume of content, which make it difficult to process without using reliable analytical techniques that can extract useful and accurate insights (Saura, Ribeiro-Soriano, & Palacios-Marqués, 2021).

One promising method for analyzing the structure of user interaction is Social Network Analysis (SNA). SNA offers a systematic approach to understanding connections among individuals and the flow of information between them. It emphasizes the importance of structural relationships over individual attributes, allowing researchers to explore how users engage and influence each other in a digital environment. This method provides an effective framework for capturing communication patterns in complex networks (Yang, Keller, & Zheng, 2020). It is also useful for identifying key actors within digital interaction structures using both theoretical and computational approaches (Rawlings, 2023).

Beyond network structures, capturing the emotional tone of user discourse is essential to understanding their experience and perception of EdTech platforms. A lexicon-based sentiment analysis approach is applied using the InSet lexicon along with contextual vocabulary. This technique evaluates the emotional tone of discourse and categorizes tweets as positive, negative, or neutral. Compared to machine learning classifiers, lexicon-based methods are more transparent and adaptable to specific language domains, making them suitable for capturing the nuances of Indonesian social media expressions in educational contexts (Homepage et al., 2021).

To further explore recurring issues and discussion themes, this study incorporates topic modelling using the BERTopic algorithm. BERTopic integrates transformer-based language embeddings with clustering techniques and class-based TF-IDF (c-TF-IDF) to generate coherent and interpretable topics from large text corpora (Grootendorst, 2022). This approach helps identify dominant concerns, interests, or campaign-related conversations, offering EdTech platforms the opportunity to align their content, support, and strategy with what matters most to their users.

Several previous studies have utilized SNA and sentiment analysis in evaluating online platforms. SNA and sentiment analysis have been applied to evaluate internet service providers through the structure of user interactions and sentiment polarity on social media (Setiadi, Mukharom, Suhendra, & Bima, 2024). In e-commerce, the combination of SNA and topic modelling has been applied to analyze user engagement patterns and identify dominant themes in online discussions (Anggraini Alamsyah & Widarmanti, 2023). A more comprehensive method, involving the integration of SNA, sentiment analysis, and topic modelling, has been used to assess public perception of fintech lending in Indonesia based on Twitter Data (Utami et al., 2022). While these studies demonstrate the value of UGC-based analysis, only a limited number have implemented all three methods, namely SNA, sentiment analysis, and topic modeling, within a single research framework. The use of BERTopic as a topic modeling technique remains limited in the Indonesian EdTech context. This gap presents an opportunity to generate more comprehensive and multidimensional insights into user interactions and perceptions.

One of the most promising approaches to address this challenge is Social Customer Relationship Management (Social CRM). Social CRM goes beyond traditional customer management by incorporating social media analytics into decision-making processes. This framework allows companies to develop more responsive, personalized, and data-driven customer engagement strategies. The implementation of a comprehensive Social CRM model that integrates interaction management, customer insight, and value co-creation has been shown to significantly improve organizational responsiveness and foster innovation in digital service environments (Rostamzadeh, Bakhnoo, Strielkowski, & Streimikiene, 2024).

This study addresses the identified research gap by conducting an integrated analysis of UGC from three major Indonesian EdTech platforms, Ruangguru, Pahamify, and Zenius. The dataset consists of tweets collected from platform X during the period of March 1-31, 2025, a strategic period in which students across Indonesia actively prepare for university entrance exams and final school assessments. This time frame was selected due to heightened user activity and increased engagement with educational platforms, making it ideal for capturing a representative sample of user discourse and sentiment. The analysis includes structural SNA metrics calculated via Gephi, sentiment polarity measured using a lexicon-based approach, and thematic topic clusters derived using BERTopic.

By combining these three methodologies, this research aims to enhance Social CRM strategies for EdTech platforms through empirical insights from social media. The results are expected to contribute both academically and practically by providing data-driven recommendations for improving digital educational services in Indonesia and supporting broader efforts to strengthen national education quality through user-centric digital innovation. Such efforts align with national priorities to strengthen inclusive and equitable digital learning ecosystems, as emphasized by UNICEF in its policy review on expanding digital education across the country.

## RESEARCH METHODS

This research uses a quantitative content analysis approach to examine social interactions, sentiment dynamics, and thematic discourse in public conversations on platform X, focusing on three Indonesian EdTech platforms, Ruangguru, Pahamify, and Zenius The dataset consists of user-generated tweets obtained via Twitter API, targeting content that reflects real-time public engagement with these platforms. The research process comprises several integrated stages, including data collection, preprocessing, social network analysis, sentiment classification, and topic modeling. Each stage was implemented using Python-based tools and Gephi for network visualization to ensure precision, transparency, and reproducibility. This methodological framework enables a multi-perspective analysis of user discourse and supports the development of data-driven strategies for Social CRM in digital education.

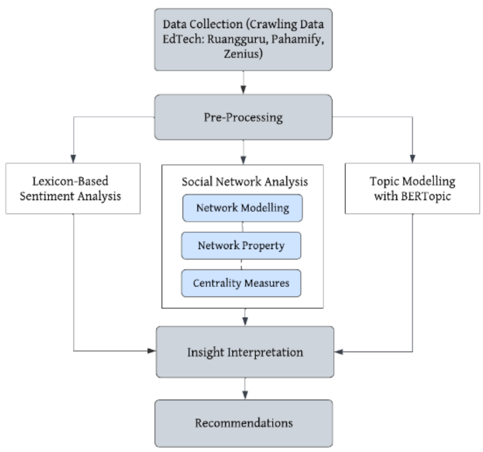


Figure 1. Research Stages

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.383
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

**Data Collection**

The data in this study were obtained from platform X through a keyword-based crawling process conducted over a defined observation period, from March 1 to March 31, 2025. The data collection focused on public tweets that mentioned or replied to official accounts of the selected EdTech platforms, or that included commonly used keywords such as “ruangguru”, “pahamify”, and “zenius”. Only tweets written in Bahasa Indonesia were retained to ensure linguistic consistency throughout the analysis. This filtering step was essential for optimizing performance during lexicon-based sentiment analysis and topic extraction.

The crawling process was conducted using the tweet-harvest tool, which was executed in a Python environment and executed via Google Collab. This tool accessed Twitter’s API v2 under academic research access using an authentication token, allowing it to retrieve filtered tweet results based on keyword, language, and time period criteria. The output of each query was saved in CSV format, which served as the primary input for the data preprocessing stage in the subsequent phase of the research.

**Pre-processing**

Data preprocessing was conducted to ensure the quality and consistency of textual data before it entered the analytical stages. The procedures performed include cleaning (removal of URLs, punctuation, emojis, and other non-linguistic elements), case folding (conversion of all characters to lowercase), normalization of informal language, tokenization, stopword removal, and stemming using the Sastrawi algorithm (Liu, Zhao, Liu, & Xu, 2015).

Text normalization and cleaning are particularly important when working with informal content like tweets, which often include misspellings, abbreviations, and emotive symbols (Murshed, Mallappa, Ghaleb, & Al-ariki, 2021). In addition, for the purposes of social network analysis, node and edge structures were extracted from user mentions and replies to construct the interaction graph. This graph served as the basis for modeling user connections in the network.

**Social Network Analysis**

Social Network Analysis (SNA) was applied in this study to understand the structure and dynamics of user interactions related to EdTech platforms on platform X. A social network is represented as a directed graph, where the set of users is modeled as nodes and the interactions between them as edges. These interactions were extracted from mentions, replies, and retweets. To evaluate the structure of the network, several network properties and centrality measures were calculated using Gephi. Gephi's implementation of centrality and network structure metrics has been widely recognized in empirical research as a reliable method for analyzing complex relational data (Vivek et al., 2025). This study focuses on 6 networks, described as follows:

a. Size

Size refers to the basic structure of the network in terms of the number of nodes and edges. Nodes represent unique users involved in interactions, while edges represent the connections established between them through mentions, replies, or retweets. A larger network size indicates higher participation and broader user involvement in the discussion around each EdTech platform.

b. Density

Density describes the degree of connectedness in the network. A higher density indicates that users in the network are more tightly connected and interact more frequently. The density value is calculated using the formula:

$$\Delta = \frac{L}{L(g-1)/2} \tag{1}$$

c. Modularity

Modularity is used to evaluate the extent to which the network is divided into communities. It reflects the strength of clustering among nodes. A high modularity value indicates well-formed community structures. Modularity is calculated using the formula:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{2}$$

d. Diameter

Diameter refers to the length of the longest shortest path between any two nodes in the network. It represents how far information must travel in the worst-case scenario. The metric is defined as:

$$D_{\max} = \max(d(i,j)) \tag{3}$$

e. Average Path Length

Average path length represents the average number of steps required to connect any two users in the network. A lower value indicates faster potential information spread. The metric is computed as:

$$\langle d\rangle = \left(\frac{1}{n(n-1)}\right)\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} d(i,j) \quad (4)$$

f. Average Degree

Average degree reflects the average number of connections or edges that each node has. This metric indicates how active the users are in the network. The formula for average degree is expressed as:

$$P_k = \frac{\sum_{i=1}^{N} k_i}{N} \quad (5)$$

In addition to evaluating the overall structural properties of the network, this study also assessed user influence and position within the network through several centrality measures. These metrics offer insight into how information flows and which users play key roles in the dissemination process. The centrality measures analyzed in this research include degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality.

a. Degree Centrality

Degree centrality is a basic centrality metric that reflects the number of direct connections a node has in the network. It indicates the level of user activity and interaction visibility. The value is computed using the formula:

$$C_D(n_i) = d(n_i) \quad (6)$$

b. Betweenness Centrality

Betweenness centrality measures how often a node appears on the shortest paths between other pairs of nodes. It represents the node's role as a bridge for information flow. The formula used is:

$$C_B(n_i) = \sum\left(\frac{g_{jk}(n_i)}{g_{jk}}\right) \quad (7)$$

c. Closeness Centrality

Closeness centrality reflects how close a node is to all other nodes in the network. It is the reciprocal of the total distance from a node to every other node. A higher closeness value means the node can quickly reach others in the network. The formula is expressed as:

$$C_C(n_i) = \frac{(N-1)}{\sum d(n_i, n_j)} \quad (8)$$

d. Eigenvector Centrality

Eigenvector centrality measures a node's influence not only based on its direct connections, but also by the importance of its neighbors. A node connected to highly influential nodes will score higher in this metric. This is useful for identifying nodes that are influential in the overall network structure.

**Lexicon-Based Sentiment Analysis**

This study employs a lexicon-based sentiment analysis method to determine the emotional tone of user comments directed toward educational technology (EdTech) platforms. Lexicon-based approaches rely on predefined dictionaries (lexicons) that contain lists of words labeled as positive, negative, or neutral. The Indonesian Sentiment Lexicon (InSet) was selected as the main reference due to its suitability for the characteristics of informal language often used on social media, and it has shown promising results when applied to real-world datasets (Asri, Kuswardani, Suliyanti, Manullang, & Ansyari, 2025).

$$S = \sum_{\{i=1\}}^{\{n\}f(w_i)}$$

Sentiment classification is performed by calculating the polarity of each tweet. This process begins with tokenizing the tweet into individual words. Each token is then compared with the lexicon. If a word exists in the positive word list, it contributes +1 to the sentiment score. If it exists in the negative word list, it contributes -1. If it is not found in either, it contributes 0. The overall sentiment of a tweet is determined by the sum of these values.

**Topic Modeling with BERTopic**

Topic modeling in this study was conducted using BERTopic to identify thematic patterns in user discourse surrounding Ruangguru, Pahamify, and Zenius on platform X. BERTopic integrates transformer-based sentence embeddings from SBERT with density-based clustering via HDBSCAN and generates topic representations using class-based TF-IDF (c-TF-IDF). Unlike traditional models such as LDA, BERTopic allows for dynamic, context-aware topic extraction, which is particularly effective for short

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v7i4.383
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

and informal text like tweets. The output consists of semantically grouped topics where each cluster is characterized by a ranked list of keywords. Topics that emerged include platform feedback, technical issues, feature requests, and promotional discussions. These thematic findings complement sentiment and network analyses by offering deeper insight into the issues that shape user engagement and public perception of EdTech services.

These three analytical approaches were selected because social media engagement in educational technology platforms involves not only user interactions (captured through network analysis) but also public perception (sentiment) and thematic focus (topics). BERTopic effectively captures nuanced themes from diverse texts, supporting its use in analyzing informal user discourse (Brawijaya, Raditya, Listyawan, Setiawan, & Saputra, 2017). This multi-perspective analysis is essential to designing effective, user-responsive Social CRM strategies in the context of Indonesian EdTech.

## RESULTS AND DISCUSSION

The initial stage of this research involved the extraction of user-generated content (UGC) from platform X related to three educational technology platforms in Indonesia, Ruangguru, Pahamify, and Zenius. The crawling process focused on public tweets that included direct mentions and replies to the official accounts of each platform, resulting in a total of 1,788 tweets for Pahamify, 907 tweets for Ruangguru, and 698 tweets for Zenius. Each tweet entry contained information such as the tweet content, user account, timestamp, and type of interaction, which was then structured into a network format. The interaction data served as the foundation for building social graphs to observe how digital communities form and engage around each EdTech service.

### Network Modelling

The interaction networks of Ruangguru, Pahamify, and Zenius were visualized using Gephi, based on directed relationships derived from mentions and replies on platform X. Each node represents a unique user, while each edge indicates a directed interaction between users. Figure 2 displays the resulting network graphs, with subfigures (a), (b), and (c) corresponding to Ruangguru, Pahamify, and Zenius, respectively. The visualizations allow for an initial examination of the conversational structure within each community. The research results are presented in graphical, tabular, or descriptive form. Analysis and interpretation of these results are required before the discussion.

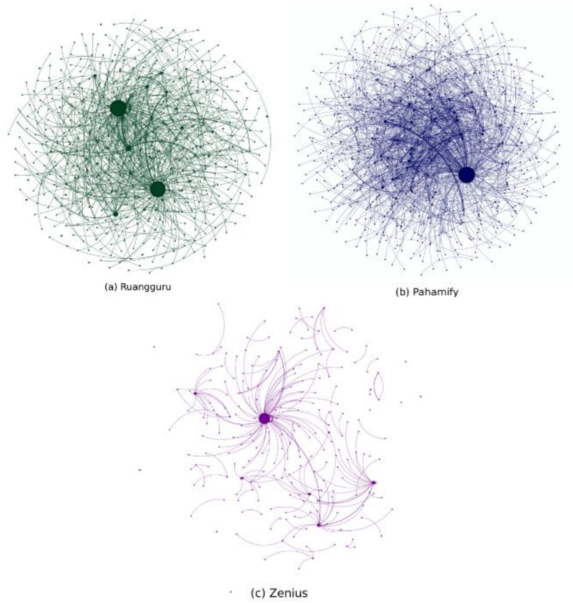


Figure 2. Social Network Visualization of (a) Ruangguru, (b) Pahamify, (c) Zenius

Ruangguru's network (Figure 2a) appears dense and centralized, with a core of highly connected users surrounded by peripheral participants. In contrast, Pahamify (Figure 2b) exhibits a more fragmented structure, suggesting a distributed conversation pattern with several disconnected clusters. Zenius (Figure 2c) presents a moderately centralized form, where interactions are evenly distributed yet still anchored by key actors. These visual characteristics provide a foundation for further structural and centrality-based analysis in the subsequent sections.

### Network Property Analysis

After conducting network modelling using Gephi, the next stage was to evaluate the structural characteristics of each network based on key graph properties. These properties provide a quantitative foundation to interpret user connectivity, community formation, and communication flow within the EdTech-related discourse. Table 1 summarizes the measurement of seven structural metrics for the Ruangguru, Pahamify, and Zenius networks. The values obtained from these properties are then analyzed to understand each platform's unique interaction behavior.

Table 1. Structural Properties of EdTech User Interaction Networks

| No. | Network Properties | Ruangguru | Pahamify | Zenius |
|---|---|---|---|---|
| 1 | Nodes | 594 | 779 | 247 |
| 2 | Edges | 522 | 991 | 214 |
| 3 | Density | 0.001 | 0.002 | 0.004 |
| 4 | Modularity | 0.876 | 0.778 | 0.745 |
| 5 | Diameter | 4 | 3 | 3 |
| 6 | Average Path Length | 1.872 | 1.141 | 1.499 |
| 7 | Average Degree | 0.879 | 1.24 | 0.866 |

The first property is the number of nodes, which indicates the total number of unique users involved in the network. The greater the number of nodes, the wider the user participation and the broader the scope of interaction. Based on Table 1, Pahamify has the largest number of nodes (779), followed by Ruangguru (594), and Zenius (247). This suggests that Pahamify attracted the highest volume of public participation during the observation period.

The second property is the number of edges, which refers to the total number of interactions between users. A higher edge count implies more dynamic conversations and a denser flow of information. Pahamify shows the highest number of edges (991), followed by Ruangguru (522) and Zenius (214), reinforcing that Pahamify's discourse was the most active and interconnected.

The third property is density, which measures how closely the nodes are connected relative to all possible connections. The higher the density, the more cohesive and tight-knit the communication structure. Zenius has the highest density value (0.004), followed by Pahamify (0.002), and Ruangguru (0.001), suggesting that despite its smaller size, Zenius users were more interconnected on average.

The fourth property is modularity, which evaluates the strength of community divisions in the network. A higher modularity score reflects clearer community segmentation and more distinct topical clusters. Ruangguru has the highest modularity value (0.876), followed by Pahamify (0.778), and Zenius (0.745), indicating that Ruangguru's discourse is the most segmented into topic-based user groups.

The fifth property is average degree, which represents the average number of direct connections each user has. The higher the average degree, the more interactions are initiated or received per user. Pahamify recorded the highest average degree (1.24), while Ruangguru and Zenius are relatively similar at 0.879 and 0.866, respectively. This implies that Pahamify's users were more communicative and engaged.

The sixth property is diameter, defined as the longest shortest path between any two nodes in the network. The smaller the diameter, the faster information can potentially reach all users. Both Pahamify and Zenius have a diameter of 3, while Ruangguru has a diameter of 4, showing that Pahamify and Zenius offer more efficient reachability across users.

The seventh property is average path length, which measures the mean number of steps required to connect any two users in the network. A lower average path length allows faster and more efficient information spread. Pahamify again leads with the shortest average path length (1.141), followed by Zenius (1.499), and Ruangguru (1.872), reinforcing its role as the most communicative network structure.

Based on the structural metrics discussed above, each EdTech platform demonstrates distinct network characteristics that reflect its user interaction dynamics. By comparing the dominance of specific properties in each platform, their respective strengths in community formation, communication flow, and network efficiency can be identified.

Ruangguru demonstrates a structured and segmented interaction pattern, characterized by well-formed clusters and distinct groupings of users. The network reflects a clear separation of topical communities, suggesting that discussions are centered around specific themes or user interests. This structure enables focused communication within each group while maintaining stable interaction across the broader network. The presence of these organized subgroups indicates that Ruangguru has a community-driven network conducive to targeted engagement and thematic discourse.

Pahamify presents an open and dynamic interaction structure with a high volume of user activity and engagement. The network shows widespread connectivity and active communication, where users frequently interact across different parts of the graph. While it may appear less segmented, the overall structure allows for rapid information exchange and high accessibility. This suggests that Pahamify facilitates broad, fast-paced discussions within a highly active user base.

Zenius represents a compact and efficient network, marked by closely linked user interactions and minimal structural distance. The network favors connectivity over segmentation, allowing users to access information quickly with few intermediaries. Despite its smaller scale, the structure remains cohesive and functionally robust. This configuration supports efficient communication and reinforces the platform's capacity to sustain active discussion within a well-connected community.

**Key Actor Identification Using Centrality Measures**

To understand the influence and strategic position of individual users in each network, four centrality metrics were calculated, there are degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. These measures help identify key actors based on how well-connected they are, how often they bridge communication paths, how quickly they can reach others, and how influential their neighbors are within the network structure. Table 2 lists the top actors from Ruangguru, Pahamify, and Zenius networks based on their centrality scores.

Table 2. Key Actors Identified Based on Centrality Measures

| Actor (Platform) | Degree | Betweenness | Closeness | Eigenvector |
|---|---|---|---|---|
| Shelovsred (Ruangguru) | 2 | 69 | 0.6 | 0.226838 |
| Yourmleux (Pahamify) | 23 | 37 | 1 | 0.017005 |
| 13434Olover (Zenius) | 67 | 131 | 0.956522 | 0.428548 |

In the Ruangguru network, the most influential actor is "shelovsred," who shows high betweenness and eigenvector centrality, despite having a relatively low degree. This indicates that the user acts as a bridge between communities, playing a strategic role in facilitating cross-cluster interactions. Although not the most active in terms of direct connections, their position within the network allows them to influence information flow significantly.

Pahamify's key actor is "yourmleux," who has the highest degree of centrality among the three platforms. This suggests that the user maintains many direct connections, making them a focal point for interaction. However, their relatively low betweenness and eigenvector values imply that their influence is more localized, operating within a specific cluster rather than across the entire network.

In contrast, the Zenius network highlights "13434Olover" as a user with consistently high scores across all centrality dimensions. The actor's high degree, closeness, and betweenness centrality indicate that they are not only well-connected but also strategically positioned to access and distribute information quickly across the network. This suggests that Zenius's discourse is strongly shaped by a single, highly influential actor who operates at the core of both communication and structure.

These findings provide insights into the structural roles of key actors in each platform, which can inform targeted engagement strategies in Social CRM. Platforms can leverage these actors either as ambassadors, information relays, or community organizers, depending on their position and type of centrality dominance.

**Lexicon-Based Sentiment Analysis**

Table 3. Sentiment Distribution Using Lexicon-Based Method on Edtech Platforms

| Sentiment Category | Ruangguru | Pahamify | Zenius |
|---|---|---|---|
| Positive | 44.6% (404) | 31.3% (556) | 41.4% (261) |
| Negative | 26% (236) | 41.8% (744) | 32.5% (205) |
| Neutral | 29.4% (266) | 26.9% (479) | 26.1% (165) |

Sentiment analysis was conducted using a lexicon-based classification approach to assess public perception of each EdTech platform. As presented in Table 3, Ruangguru received the highest proportion of positive sentiment (44.6%), indicating generally favorable user responses, followed by Zenius (41.4%). Pahamify, in contrast, exhibited the highest percentage of negative sentiment (41.8%), suggesting a relatively more critical or dissatisfied audience. Neutral sentiment was most evenly distributed across platforms, with values ranging between 26% and 29%, indicating that a substantial portion of the conversation remained objective or informational.

**Topic Modelling with BERTopic**

To identify the dominant themes discussed by users, topic modelling was applied using the BERTopic algorithm, which combines sentence embeddings with clustering and class-based TF-IDF weighting. This method enabled the categorization of user conversations into meaningful topic groups based on semantic similarity. Each EdTech platform exhibited different thematic focuses, reflecting varying user concerns and interests. Topics ranged from learning features, pricing models, and technical issues to platform promotions and user feedback.

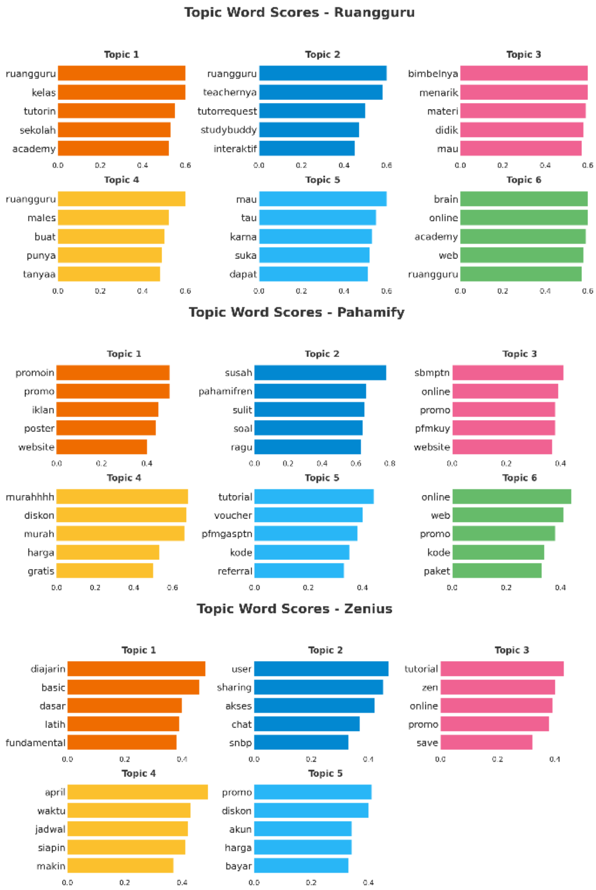


Figure 3. Topic Clusters by BERTopic for Ruangguru, Pahamify, and Zenius

Across the platforms, Ruangguru's discussions were dominated by references to product updates, promotional campaigns, and interactive learning content. Pahamify's topic clusters not only reflect promotional vocabulary, discount codes, and exam-related offerings, but also include several terms that indicate user concerns, such as difficulties in navigation or learning comprehension. In contrast, Zenius generated a more balanced theme distribution, encompassing both platform appreciation and requests for content improvements. These thematic distinctions serve as valuable input for platform managers to tailor CRM strategies according to the specific needs and expectations voiced within each digital community.

In relation to sentiment distribution, the themes extracted through BERTopic also reflect the emotional tone found in the user discourse of each platform. For instance, clusters related to "susah," "ragu," and "sulit" appeared in discussions about Pahamify, which corresponded with a higher proportion of negative sentiment in the sentiment analysis phase. Despite this, several other topic clusters on Pahamify also contained references to product promotions and exam-related content, suggesting a mix of both critical feedback and service-oriented discussion. Ruangguru exhibited topic groups centered on "exam preparation" and "motivational content," aligning with its larger share of positive user sentiment. Zenius displayed more neutral themes involving "content structure" and "lesson format," consistent with its balanced sentiment polarity. These correlations between topic composition and sentiment polarity suggest that each platform's public discourse is shaped not only by service quality but also by how users emotionally respond to specific features or issues.

Specifically for Zenius, the topic modeling process using BERTopic resulted in only five valid topic clusters. This limitation occurred due to the relatively lower number of tweets collected from this platform compared to Ruangguru and Pahamify. In addition, the text data from Zenius users exhibited reduced lexical diversity, causing several candidate topics to merge or be excluded during the clustering stage. Despite this, the identified themes, such as "learning structure" and "exam readiness," still provided valuable insight into user focus areas and platform perception.

## Insight Interpretation

The combined results of social network analysis, sentiment classification, and topic modeling provide a comprehensive perspective on user interaction patterns, perception, and thematic focus surrounding Indonesian EdTech platforms. Each method captures a distinct dimension. SNA highlights structural relationships and key actors, sentiment analysis indicates the tone of user responses, while topic modeling exposes the thematic direction of public discourse. The triangulation of these methods allows for a deeper understanding of how users engage with Ruangguru, Pahamify, and Zenius on platform X, not only in terms of volume but also intensity and content quality.

Ruangguru demonstrates a structured and community-oriented network, supported by clear modular segmentation and a high degree of clustering. The prevalence of positive sentiment aligns with frequently mentioned themes such as promotional campaigns, motivational content, and interactive learning features. These findings indicate that Ruangguru has succeeded in establishing brand trust and sustained user engagement, especially through centralized communication and content-driven appeal. The presence of strong core actors reinforces its stability in navigating online discourse.

Pahamify, in comparison, reflects an active interaction network with a relatively less centralized structure. The sentiment analysis indicates a greater presence of negative expressions, with topic clusters containing terms such as "susah," "sulit," and "ragu." These discussions likely represent user feedback concerning perceived challenges in navigating the platform. Although these expressions do not necessarily reflect widespread dissatisfaction, they highlight areas where service improvements may be needed. Pahamify's active user engagement presents an opportunity for the platform to better understand user concerns and implement more responsive support mechanisms aimed at enhancing overall satisfaction.

Zenius, although smaller in scale, presents a cohesive network with high density and relatively balanced sentiment distribution. User discussions encompass both appreciation for content and constructive suggestions for improvement. While the platform does not lead in engagement volume, its efficient communication structure and moderate emotional tone suggest a healthy discourse environment. Zenius's interaction model may offer resilience and consistency, especially in niche communities with focused educational needs. These attributes position Zenius as a promising platform for delivering targeted educational services, particularly for learners seeking structured and self-paced experiences.

**Strategic Social CRM Recommendations**

Building on the results from the structural, sentiment, and thematic analyses, this section proposes platform-specific Social CRM strategies. Figure 4 provides a visual summary of these recommendations, highlighting how each EdTech platform can tailor its engagement approach based on its user interaction pattern and discourse dynamics.

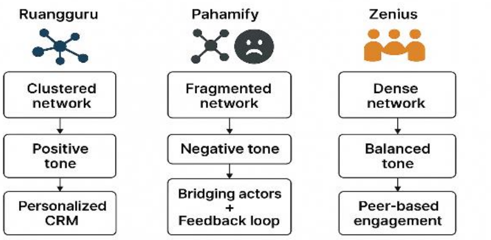


Figure 4. Social CRM Mapping for Ruangguru, Pahamify, and Zenius

Ruangguru is encouraged to continue strengthening its Social CRM approach by leveraging its well-formed community structure and centralized interaction hubs. The modular nature of its user network presents an opportunity to deliver more personalized content and targeted engagement aligned with the specific interests of each cluster. To maintain positive sentiment, Ruangguru should implement proactive sentiment monitoring and community listening mechanisms, allowing the platform to respond quickly to emerging concerns. This approach aligns with the platform's high modularity score (0.876), indicating clear thematic segmentation across user clusters. Additionally, key users such as "shelovsred," who scored high in both betweenness and eigenvector centrality, could serve as effective community ambassadors to reinforce targeted communication strategies. This reinforces user trust and supports long-term relationship management within a stable digital community.

Pahamify, which displays fragmented interaction patterns and a relatively higher volume of negative sentiment, may benefit from strategies aimed at rebuilding user trust. Responsive communication and transparent service updates are essential to address areas perceived as challenging, particularly those reflected in topic clusters mentioning difficulties and uncertainty. With a lower modularity score (0.415), the platform's network appears less cohesive, suggesting the need to strengthen inter-user connections. From a Social CRM perspective, activating bridging users such as "yourmleux" could help connect disparate user segments and facilitate a more unified conversation. Establishing feedback channels and acknowledging user voices in public spaces may further assist in reframing perceptions and supporting brand recovery efforts.

Zenius, although operating on a smaller interaction scale, demonstrates strong cohesion and a balanced emotional tone. The platform can build upon these strengths by nurturing peer-to-peer learning networks and promoting organic user initiatives. Its high graph density and average degree metrics suggest that users are already interconnected, enabling efficient information flow within the network. Key actors such as "13434olover," who demonstrates high centrality across degree, betweenness, and eigenvector measures, may be engaged to facilitate community interaction or moderate knowledge-sharing sessions. Sustaining this collaborative interaction model will allow Zenius to deepen user engagement in niche communities while maintaining user loyalty.

This community-centric strategy aligns with Social CRM principles by fostering collaborative relationships and amplifying the user voice in shaping platform direction, as has been demonstrated in other domains where relationship marketing and customer focus improve organizational performance (Letchumannan, Bidin, Bolong, & Osman, 2022). Aligning with Shalihati, Sumarwan, Hartoyo, & Yuliati (2025), the integration of AI-based insights and multichannel strategies into Social CRM is expected to improve educational service delivery and institutional responsiveness. These principles, if implemented consistently, could significantly enhance the adaptability of EdTech platforms in aligning with evolving user expectations in digital learning.

## CONCLUSIONS AND SUGGESTIONS

### Conclusion

This study demonstrated the relevance of combining Social Network Analysis, lexicon-based sentiment classification, and BERTopic modeling to extract structural, emotional, and thematic insights from user-generated content related to Indonesian EdTech platforms. The analysis revealed distinct interaction patterns, sentiment orientations, and discussion topics across Ruangguru, Pahamify, and Zenius, reflecting each platform's responsiveness and community engagement. These findings are not only valuable for EdTech developers seeking to enhance their Social CRM strategies but also meaningful for the public by offering clearer perspectives on how learning platforms are experienced and discussed online. This transparency helps users make informed choices and reinforces their role in shaping the direction of digital learning services. Overall, this research contributes to strengthening the education ecosystem in Indonesia through more inclusive, responsive, and community-driven innovation.

### Suggestion

Future research should consider extending the observation period to capture long-term patterns and temporal shifts in user sentiment and interaction. A longitudinal analysis could provide deeper insights into how user behavior evolves in response to platform changes or educational trends. In addition, the application of deep learning-based models such as LSTM or transformer variants may improve the accuracy of sentiment classification and topic interpretation, especially in handling informal, context-rich social media data. These enhancements would support more granular and adaptive analysis, enabling EdTech platforms to respond with greater precision to user needs and discourse dynamics.

## REFERENCES

Anggraini Alamsyah, R., & Widarmanti, T. (2023). Improving Customer Education: Study Of Customer Engagement Of Tokopedia And Shopee On Twitter Using Social Network Analysis. *International Journal of Educational Management and Innovation*, *4*(2), 74–85. https://doi.org/10.12928/ijemi.v4i2.7728

Asri, Y., Kuswardani, D., Suliyanti, W. N., Manullang, Y. O., & Ansyari, A. R. (2025). Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application. *Indonesian Journal of Electrical Engineering and Computer Science*, *38*(1), 677. https://doi.org/10.11591/ijeecs.v38.i1.pp677-688

Brawijaya, U., Raditya, B., Listyawan, N., Setiawan, N. Y., & Saputra, M. C. (2017). *Fakultas Ilmu Komputer Topic Modelling Pada Aktivitas Pengembangan Perangkat Lunak Menggunakan BERTopic* (Vol. 1). Retrieved from http://j-ptiik.ub.ac.id

Chadwick, S., Parry, K. D., Ahmed, W., & Fenton, A. (2023). Understanding sporting brands and entrepreneurship using netnography and social network analysis. *International Journal of Technology Transfer and Commercialisation*, *20*(1), 141. https://doi.org/10.1504/ijttc.2023.10054827

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. Retrieved from http://arxiv.org/abs/2203.05794

Homepage, J., Musfiroh, D., Khaira, U., Eko, P., Utomo, P., Suratno, T., … Teknologi, D. (2021). MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon, *1*, 24–33.

Kadek, I., Divananda, D. A., & Rubiyanti, N. (2024). *Social Media Marketing and Purchase Intention of EdTech Platforms in Indonesia: The Role of Consumer Brand Engagement*. *Journal of Information Systems Engineering and Management* (Vol. 2025). Retrieved from https://www.jisem-journal.com/

Komljenovic, J., Birch, K., & Sellar, S. (2024). Monetising Digital Data in Higher Education: Analysing the Strategies and Struggles of EdTech Startups. *Postdigital Science and Education*. https://doi.org/10.1007/s42438-024-00505-0

Letchumannan, R., Bidin, R., Bolong, J., & Osman, M. N. (2022). Social Customer Relationship Management Practices And Organizational Performance In Public Sector: An Exploratory Study. *International Journal of Academic Research in Business and Social Sciences*, *12*(5). https://doi.org/10.6007/ijarbss/v12-i5/13881

Liu, B., Zhao, J., Liu, K., & Xu, L. (2015). Book Review Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. *Press*.

Murshed, B. A. H., Mallappa, S., Ghaleb, O. A. M., & Al-ariki, H. D. E. (2021). Efficient Twitter Data Cleansing Model for Data Analysis of the Pandemic Tweets. In *Studies in Systems, Decision and Control* (Vol. 348, pp. 93–114). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-67716-9_7

Rawlings, C. M. S. J. A. M. J. M. D. A. (2023). *Network Analysis: Integrating Social Network Theory, Method, and Application with R.*

Rostamzadeh, R., Bakhnoo, M., Strielkowski, W., & Streimikiene, D. (2024). Providing an innovative model for social customer relationship management: Meta synthesis approach. *Journal of Innovation and Knowledge*, *9*(3). https://doi.org/10.1016/j.jik.2024.100506

Sabilla, K., & Hartarto, R. B. (2024). A social media analysis of students' perspective toward online learning during the COVID-19 pandemic. *Cogent Social Sciences*, *10*(1). https://doi.org/10.1080/23311886.2024.2376856

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021, October 1). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*. Elsevier Ltd. https://doi.org/10.1016/j.ijinfomgt.2021.102331

Setiadi, T., Mukharom, G., Suhendra, B., & Bima, S. (2024). Application of Social Network Analysis for Comparison and Ranking of Internet Service Providers. *Jurnal Riset Informatika*, *6*(2), 77–84. https://doi.org/10.34288/jri.v6i2.280

Shalihati, F., Sumarwan, U., Hartoyo, H., & Yuliati, L. N. (2025). Mapping Customer Relationship Management Research in Higher Education: Trends and Future Directions. *Administrative Sciences*, *15*(2). https://doi.org/10.3390/admsci15020068

Utami, H., Purnama, A. A., Hidayanto, A. N., Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). *Fintech Lending in Indonesia: A Sentiment Analysis, Topic Modelling, and Social Network Analysis using Twitter Data*. *Netherland International Journal of Applied Engineering & Technology Copyrights @ Roman Science Publications* (Vol. 4).

Vivek, N., Clark, E., Gao, L., Xu, S., Baskauf, S., Nguyen, K., … Gelbard, A. (2025). Social network analysis as a new tool to measure academic impact of physicians. *Laryngoscope Investigative Otolaryngology*, *10*(1). https://doi.org/10.1002/lio2.70060

Yang, S., Keller, F. B., & Zheng, L. (2020). *Social Network Analysis: Methods and Examples*. *Social Network Analysis: Methods and Examples*. SAGE Publications, Inc. https://doi.org/10.4135/9781071802847