Sentiment Analysis of Roblox App Reviews: Correlating User Feedback with Ratings Using Lexicon and Machine Learning Methods

Shuang Li[1,*], Matee Pigultong[2]

[1]Vocational Education Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani,12110, Thailand

[2]Educational Technology and Communications Division, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, 12110, Thailand

ABSTRACT

The study presents a detailed sentiment analysis of user feedback for the Roblox app, focusing on correlating review sentiments with numerical ratings. Utilizing both lexicon-based and machine learning techniques, the research examined 320,000 reviews from the Google Play Store. The VADER lexicon-based analysis classified approximately 76% of reviews as positive, 18% as negative, and 6% as neutral. This distribution reflected a predominantly positive sentiment, aligning with an overall trend of user satisfaction. Pearson and Spearman correlation coefficients were calculated to evaluate the relationship between sentiment scores and user ratings, yielding moderate positive relationships with values of 0.47 and 0.36, respectively. These correlations indicated that, while sentiment scores generally paralleled user ratings, other factors likely influenced rating variations. In the machine learning analysis, the study utilized Naive Bayes and Logistic Regression models for sentiment classification. Logistic Regression achieved a slightly higher accuracy of 86.4% compared to Naive Bayes' 85.3%, showing improved precision and recall for positive and negative sentiments. However, both models struggled with the neutral category, reflecting challenges in sentiment differentiation. Cross-validation confirmed the stability of the models, with the Logistic Regression model maintaining a consistent accuracy across folds. These findings underscore the potential of sentiment analysis in providing actionable insights for developers, highlighting specific areas where user sentiment diverges from ratings. Future research could enhance sentiment detection accuracy by incorporating advanced deep learning models and extending the analysis to include data from other platforms, offering a more comprehensive understanding of user feedback across gaming ecosystems.

Keywords Sentiment Analysis, Roblox App Reviews, Machine Learning, User Feedback, Correlation Analysis

Submitted 20 February 2025
Accepted 20 May 2025
Published 16 September 2025

Corresponding author
Shuang Li,
l.shuang@rmutt.ac.th

Additional Information and Declarations can be found on page 320

DOI: 10.47738/jdmdc.v2i3.38

Introduction

The rise of mobile gaming has revolutionized the global gaming industry, especially through widely accessible platforms like the Google Play Store. As smartphones become ubiquitous and mobile internet connections improve, mobile games have become one of the most engaging forms of entertainment for millions worldwide. The growth of this industry has been driven by the adoption of freemium models, where games are free to download but generate revenue through in-app purchases and ads [1]. This model has significantly lowered the entry barrier, enabling users to access and engage with games

How to cite this article: S. Li, M. Pigultong, "Sentiment Analysis of Roblox App Reviews: Correlating User Feedback with Ratings Using Lexicon and Machine Learning Methods," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 3, pp. 298-322, 2025.

without upfront costs. Games like "Mobile Legends" and "Arena of Valor" exemplify this approach, attracting millions of players and contributing to the rise of mobile gaming as a dominant force in the entertainment sector.

Moreover, mobile gaming's widespread adoption has also been influenced by shifting consumer demographics. The demographic profile of mobile gamers is diverse, with a notable increase in adolescent and younger users, a group that has shown high engagement with mobile games. The integration of social and competitive elements within these games has further contributed to their popularity, as these features encourage user retention and prolonged engagement [2]. The COVID-19 pandemic, which limited physical interactions, further accelerated the growth of mobile gaming as users turned to mobile platforms for social connection and entertainment [3]. This period of heightened demand has led to rapid innovation and a broader range of game offerings available on platforms like the Google Play Store [4].

User reviews are essential in determining the success and longevity of mobile games, particularly on platforms like the Google Play Store. These reviews act as a real-time feedback loop, offering developers insights into player experiences, expectations, and issues. Mobile games rely heavily on user engagement and satisfaction to maintain their player base, and reviews provide crucial data for developers to evaluate game performance and areas of improvement. Analyzing user reviews allows developers to understand common concerns, identify emerging trends, and make informed decisions about game updates, features, and bug fixes. This feedback gathered through both quantitative ratings and qualitative textual analysis, is integral in refining the user experience and maintaining player retention over time [5].

In addition to shaping game development, user reviews significantly impact a game's reputation and its ability to attract new players. Positive reviews contribute to the game's perceived quality, often driving more downloads and creating a virtuous cycle of engagement. On the other hand, negative reviews can serve as early indicators of declining player satisfaction, prompting necessary interventions. In a competitive mobile gaming market with thousands of options, reviews often serve as a critical differentiator that can make or break a game’s success. Studies in big data text mining, such as those by [6], demonstrate how collective sentiment derived from user feedback influences the broader perception of a game and can guide future development strategies to align with player preferences.

Moreover, user reviews serve not only as feedback for developers but also as a means of fostering community and social interaction among players. Online discussions, often initiated in review sections, encourage community building, which is crucial for games that rely on long-term engagement and multiplayer dynamics. According to [7], these interactions increase user viscosity, or loyalty, as players feel a stronger connection to the game and each other. This sense of community, enhanced by the exchange of feedback, plays a pivotal role in a game’s long-term retention and overall success, as it deepens the emotional investment of the players. Thus, user reviews are not just passive reflections of player satisfaction but active components in building a thriving, engaged user base.

Roblox has become one of the most popular games on the Google Play Store, amassing over 300,000 reviews and having a global user base that is predominantly composed of children and teenagers. Its unique platform allows users to create and share their games and experiences, setting it apart from traditional gaming models. This user-generated content (UGC) system has propelled Roblox to the forefront of mobile gaming, as it taps into players' creativity and fosters a highly engaging environment where users can both play and develop games. According to [8], Roblox now hosts over 55 million games, spanning a wide variety of genres and interests, making it a versatile and inclusive platform that appeals to a broad audience.

One of the primary reasons for Roblox's enduring popularity lies in the strong community of players and developers it cultivates. The platform encourages collaboration by allowing users to create avatars, participate in multiplayer games, and interact socially within game environments. This collaborative aspect, combined with the ability to design custom games, gives players a sense of ownership and involvement in the game’s ecosystem [9]. The ability to contribute to the platform not only enhances the gaming experience but also strengthens player loyalty. This, in turn, creates a self-sustaining cycle where player engagement fuels the creation of new content, which attracts more users and keeps the platform dynamic and ever-expanding.

Moreover, Roblox's educational potential has been increasingly recognized, with the platform being used as a tool to enhance problem-solving skills and digital literacy among students. Studies suggest that engaging with Roblox helps foster critical thinking, programming knowledge, and even multimodal literacy [10]. This educational dimension has expanded Roblox’s appeal beyond entertainment, positioning it as a valuable resource in schools and for parents seeking enriching activities for their children. The platform's integration of educational features, such as teacher accounts and classroom management tools, underscores its versatility, offering a blend of learning and play [11]. While Roblox continues to grow in popularity, it faces challenges such as ensuring the safety and appropriateness of content, especially for its young user base [12].

Analyzing large volumes of user reviews presents a significant challenge for game developers, especially in the context of popular platforms like Roblox. With over 300,000 reviews on the Google Play Store alone, manually reviewing and understanding user feedback takes time and effort. These textual reviews contain valuable information about user experiences, preferences, and frustrations, but extracting actionable insights from them is difficult without automation. Text-based analysis becomes increasingly complex when combined with the need to quantify sentiment and correlate it with numerical ratings provided by users. This lack of an efficient system to interpret and process vast amounts of user-generated content leaves a gap in understanding how textual sentiment reflects user ratings.

Additionally, there is a notable gap in the literature regarding how textual sentiment correlates with numerical ratings in mobile game reviews, particularly for large datasets like that of Roblox. Existing studies on sentiment analysis within the gaming sector often focus on smaller datasets or different contexts, leaving a need for more comprehensive research in this area. While sentiment analysis has been applied to product reviews and social media posts, few studies address how user feedback for games can be systematically analyzed to guide development decisions. This research, therefore, aims to fill that gap by examining the relationship between review text and user ratings, contributing to a deeper understanding of how sentiment is expressed in game reviews.

The primary goal of this study is to perform sentiment analysis on Roblox user reviews, with the aim of correlating text-based sentiment with numerical ratings. The research focuses on using lexicon-based and machine learning methods to extract sentiment from review text and then analyze the connection between these sentiment scores and the ratings users provide. By achieving this, the study aims to uncover patterns in user feedback that can help developers better understand how players perceive their game.

To accomplish this, the study has three key objectives: first, to extract sentiment from user reviews using both lexicon-based and machine learning techniques; second, to analyze the correlation between these sentiment scores and the associated user ratings; and third, to identify discrepancies between the expressed sentiment in the text and the numerical ratings. These discrepancies can reveal important insights into user dissatisfaction or misunderstood features, helping developers refine their games to meet player expectations better.

This research is significant as it contributes to the gaming industry by providing a structured approach to understanding user feedback at scale. The ability to systematically analyze vast amounts of user reviews enables developers to gain insights into how players perceive game updates, features, and overall performance. By correlating sentiment with numerical ratings, this study not only improves the understanding of user feedback but also helps developers prioritize areas that need improvement, such as gameplay mechanics or user interface design.

Furthermore, this study’s findings can guide future game development by offering a clearer picture of user preferences and frustrations. For a game like Roblox, where user-generated content and regular updates are critical to its success, the insights gained from sentiment analysis can inform better decision-making in terms of feature development and community engagement. By leveraging these insights, developers can enhance the overall player experience and ensure sustained engagement with the game, making this research valuable for both the gaming industry and the broader field of sentiment analysis.

Literature Review

Sentiment Analysis in Gaming

Sentiment analysis (SA) has gained substantial relevance in the gaming industry, as it enables the extraction of player sentiments from vast amounts of user-generated content. As the gaming sector continues to expand, players frequently share their opinions on gaming platforms, forums, and social media, making sentiment analysis an essential tool for understanding these diverse perspectives. The primary function of sentiment analysis in gaming is to assess player reviews, social media posts, and other forms of feedback to identify the general public's attitude toward a game. This analysis provides actionable insights for developers, particularly in improving game design, implementing updates, and tailoring marketing strategies [13].

Recent studies have underscored the significance of sentiment analysis and machine learning in understanding user feedback and behavior across various digital platforms, aligning closely with the goals of this study on Roblox app reviews. For instance, research on the myIM3 mobile application explored complex emotional landscapes, revealing that user sentiment analysis can capture multifaceted emotional expressions and provide insights into app satisfaction and areas for improvement [14]. Similarly, an analytical study of Amazon product reviews examined how discount strategies influence consumer ratings, demonstrating how external factors can impact sentiment and ratings—an observation that is also relevant for understanding discrepancies in user feedback for the Roblox app [15]. Moreover, a comparative analysis of sentiment classification techniques on Flipkart product reviews, employing models like Logistic Regression and Random Forest, highlights the effectiveness of machine learning in categorizing user sentiment, which is integral to this study's approach of correlating sentiment with ratings in mobile app reviews [16]. Additionally, customer segmentation using Gaussian Mixture Models in digital advertising underscores the potential of advanced clustering methods to group users based on behavioral patterns, a technique that could further enhance insights gained from sentiment analysis in app reviews [17].

Other studies in the domain of sentiment analysis and behavioral trends offer methodological insights relevant to analyzing Roblox app reviews. For example, research on Bitcoin-related tweets used TF-IDF vectorization and K-Means clustering to identify sentiment trends, underscoring the value of feature extraction and clustering techniques in understanding public perception, which parallels this study’s text processing approach [18]. Additionally, an analysis of the correlation between Bitcoin trading volume and price movements using Pearson and Spearman methods demonstrates the utility of correlation analysis in examining relationships between user behavior and quantitative metrics, mirroring this study’s focus on correlating sentiment scores with review ratings [19]. In a study focusing on the metaverse, predictive modeling of Roblox stock prices using machine learning and time series analysis emphasizes the role of predictive models in identifying trends and patterns within digital platforms, reflecting the potential for similar techniques in analyzing user feedback trends in mobile app ecosystems [20]. Finally, a comprehensive analysis of Twitter conversations related to the metaverse offers insights into discourse patterns that inform sentiment trends, providing a broader perspective on user engagement that can be applied to analyze feedback trends in the Roblox app review dataset [21]. These studies collectively demonstrate the versatility of sentiment analysis and machine learning techniques, validating their application to app reviews as a means of understanding user satisfaction, identifying potential improvements, and uncovering the drivers behind user ratings.

A significant challenge in applying sentiment analysis to gaming is the complexity of language used in game reviews. Game-related discussions often involve sarcasm, humor, and domain-specific terminology, complicating the interpretation of textual sentiment. Research [13] emphasizes that conventional sentiment analysis tools may misinterpret these linguistic features, resulting in inaccurate sentiment classification. To address these challenges, recent studies have focused on developing advanced natural language processing (NLP) models tailored to the gaming domain. Research [22] argue that these domain-specific models are crucial for capturing the nuances of gaming-related language, enabling more accurate sentiment classification. The incorporation of these refined models allows developers to understand the better player sentiment and its correlation with user satisfaction and game success.

Sentiment analysis plays an instrumental role in improving game development by providing insights into how players perceive specific game features. Aspect-based sentiment analysis (ABSA) is particularly effective in this regard, as it allows developers to break down player feedback into distinct aspects, such as gameplay, graphics, and narrative. By focusing on these detailed aspects, developers can prioritize design enhancements that directly align with player expectations. Additionally, sentiment analysis supports the iterative nature of game development. Continuous monitoring of player feedback through sentiment analysis enables developers to implement real-time updates based on player sentiment. [23] highlight how game developers can adjust their strategies based on user reviews and sentiments, leading to a more agile and responsive development process. This feedback loop, which involves identifying player sentiments and translating them into actionable changes, enhances user engagement and satisfaction. Sentiment analysis thus serves as a feedback mechanism that not only improves game features but also aligns the game's evolution with player expectations.

Beyond improving game design, sentiment analysis plays a vital role in gauging player satisfaction and loyalty. By analyzing the emotional tone of player feedback, developers can better understand the overall player experience and implement measures to enhance retention. Research by [24], although centered on sentiment analysis within social media platforms, provides methodologies that can be adapted for gaming contexts to evaluate user engagement. Sentiment analysis helps developers track emotional responses to game updates or changes, identifying potential dissatisfaction early and addressing it before it leads to higher churn rates.

Moreover, sentiment analysis is increasingly utilized for quality assurance by identifying recurring issues faced by players. Study [25] explore how feedback from user reviews can be analyzed to highlight common bugs or frustrations, allowing developers to refine their games pre-release or through patches. The ability to detect these issues early on through sentiment analysis significantly improves the overall player experience, contributing to the game's longevity. Lastly, sentiment analysis also informs marketing strategies by highlighting aspects of the game that resonate most positively with players. Research [26] emphasize the importance of leveraging sentiment data to craft more effective marketing campaigns, ensuring that developers can better target their audience and enhance player acquisition and retention efforts.

Lexicon-Based Sentiment Analysis

Lexicon-based sentiment analysis is a widely utilized technique in natural language processing (NLP) that assigns sentiment scores to words based on predefined dictionaries. This approach classifies words as positive, negative, or neutral, often accompanied by intensity scores to indicate the degree of sentiment conveyed. The fundamental concept behind lexicon-based methods is to match words in a given text against a lexicon—essentially a sentiment dictionary—and calculate an aggregate sentiment score for the text [27], [28]. This methodology is particularly effective in analyzing structured text or contexts where sentiment-carrying words can be easily identified and quantified [29].

The creation of sentiment lexicons follows either manual or automated approaches. Manual lexicons are typically developed by linguistic experts who annotate words according to their semantic orientation, while automated methods derive sentiment scores from vast corpora using algorithms. For example, the SentiWordNet lexicon is constructed by assigning sentiment scores to words within the WordNet database, making it a powerful tool for many NLP tasks [30]. Additionally, hybrid methods that combine lexicon-based approaches with machine learning techniques offer improved accuracy by integrating contextual information from the text, enabling a more nuanced interpretation of sentiment [31], [32]. While lexicon-based sentiment analysis offers clear advantages, such as ease of implementation and interpretability, it is not without challenges, particularly when dealing with slang, sarcasm, or domain-specific language, where the static nature of lexicons may lead to misinterpretation [33].

Several lexicon-based tools, including VADER (Valence Aware Dictionary and Sentiment Reasoner), TextBlob, and AFINN, have gained prominence for their ability to analyze sentiment efficiently. VADER is specifically designed to analyze social media text, where informal language and brevity are common. VADER's lexicon incorporates not only traditional sentiment words but also accounts for contextual nuances, such as the effect of punctuation, capitalization, and even emojis, enhancing its accuracy in informal settings like social media posts and short reviews [34]. In fact, studies have shown VADER to achieve an F1 score of 0.96 in classifying sentiment on Twitter, outperforming many traditional machine learning models [35].

TextBlob, another commonly used tool, provides a broader framework for natural language processing tasks, including sentiment analysis. It utilizes a rule-based approach to assign sentiment polarity and subjectivity scores to text. TextBlob is especially advantageous for developers due to its simplicity and ability to handle both formal and informal text [36], [37]. On the other hand, AFINN operates on a simpler, lexicon-based system, assigning positive or negative values to words and calculating a cumulative sentiment score for a given text. Though less complex than VADER or TextBlob, AFINN's straightforwardness makes it a valuable tool for quickly assessing sentiment in various domains, including social media and product reviews [38]. Each of these tools offers distinct advantages depending on the text's context and the desired depth of sentiment interpretation.

Lexicon-based sentiment analysis tools have proven to be highly effective in analyzing mobile game reviews, offering developers quick insights into user sentiment. VADER, with its proficiency in handling informal language, has been particularly useful in analyzing reviews from mobile gaming platforms, where user feedback often includes slang, abbreviations, and emojis [39]. This capability allows developers to gain a nuanced understanding of player emotions without requiring extensive preprocessing or manual tagging of text, making the tool both efficient and scalable for large datasets [39]. VADER’s ability to interpret sentiment in this manner enables developers to quickly identify areas for improvement in gameplay or features, improving user satisfaction.

TextBlob also holds significant value in mobile game sentiment analysis due to its ability to calculate both polarity and subjectivity. The polarity score reflects the overall sentiment, while subjectivity indicates the extent to which the review reflects personal opinions rather than factual information [40]. This dual capability is particularly useful for developers when assessing player feedback, as it helps distinguish between highly opinionated reviews and those based on objective experiences. Meanwhile, AFINN's simplicity allows for rapid deployment, making it a suitable choice for analyzing large volumes of mobile game reviews when developers need quick, interpretable sentiment data [41]. In summary, these tools offer a combination of speed, interpretability, and adaptability, enabling mobile game developers to efficiently assess user sentiment and make data-driven decisions for enhancing game features and overall player experience.

Machine Learning-Based Sentiment Analysis

Supervised learning models are integral to sentiment classification, providing a systematic approach to categorizing text based on the emotions conveyed. Among the most widely used models for sentiment analysis are Naive Bayes, Support Vector Machines (SVM), and Logistic Regression. These models excel in identifying patterns within textual data, allowing for the accurate classification of sentiments expressed in reviews or social media posts.

Naive Bayes is a probabilistic classifier that operates on the assumption of feature independence. Despite this assumption, Naive Bayes is known for its simplicity and efficiency, making it a popular choice for sentiment analysis tasks involving large datasets [42]. Studies have demonstrated its effectiveness in various domains, from analyzing movie reviews to classifying political news [43], [44]. However, its assumption of independence between features can be a limitation in more complex sentiment expressions, where context and word relationships play a crucial role [45].

Support Vector Machines (SVM), on the other hand, are more robust for handling high-dimensional data like text. SVM works by finding a hyperplane that best separates different sentiment classes in the feature space. Its performance is highly dependent on kernel selection and hyperparameter tuning, which makes it both flexible and effective in sentiment classification tasks, especially when dealing with nuanced sentiments [46]. Studies have shown that SVM often outperforms Naive Bayes, particularly in tasks involving complex text, such as tweets and online reviews [47]. Lastly, Logistic Regression is frequently employed for binary classification tasks in sentiment analysis. It models the probability that a text belongs to a specific sentiment category and is valued for its simplicity and interpretability. When combined with effective feature extraction techniques like TF-IDF, Logistic Regression performs comparably to more complex models like SVM [48].

Feature extraction is critical in converting unstructured text into a format suitable for machine learning models. Without this transformation, algorithms cannot process and classify text data. Two commonly used feature extraction methods in sentiment analysis are Term Frequency-Inverse Document Frequency (TF-IDF) and n-grams. Both methods enhance the ability of machine learning models to capture the intricacies of textual data.

TF-IDF is a numerical representation of a word's importance within a document relative to a larger corpus. By considering both the frequency of a word in a specific document and its prevalence across other documents, TF-IDF reduces the impact of common words that contribute little to understanding sentiment. Research has shown that TF-IDF significantly improves the performance of classifiers such as SVM and Logistic Regression by focusing on the most relevant words in a text [48]. Additionally, TF-IDF helps mitigate the problem of high dimensionality in text classification, as it reduces the number of features while maintaining the text's meaningful content [49].

N-grams, which capture sequences of words (or characters), are another powerful feature extraction technique. By accounting for word order and context, n-grams preserve syntactic relationships that are often lost in simpler models like Bag of Words (BoW). Bigrams (two-word sequences) and trigrams (three-word sequences) are particularly useful in sentiment analysis, where the combination of words can drastically change the meaning [50], [51]. Studies indicate that models utilizing n-grams often outperform those relying solely on unigrams, as they better capture the nuances of language in tasks like sentiment classification [52]. The integration of TF-IDF and n-grams in feature extraction enhances the model's ability to interpret text, providing a more robust foundation for accurate sentiment analysis.

Evaluating the performance of machine learning models in sentiment analysis involves several key metrics: Accuracy, Precision, Recall, and F1-Score. Each of these metrics offers unique insights into a model's strengths and limitations, enabling researchers to assess the model's overall effectiveness.

Accuracy measures the proportion of correctly classified instances out of the total number of instances, providing a general sense of how well the model performs. While accuracy is a widely used metric, it can be misleading in imbalanced datasets where one class dominates. In such cases, a model might achieve high accuracy by predicting the majority class yet fail to correctly identify the minority class [53]. Precision focuses on the accuracy of positive predictions, reflecting the proportion of true positive instances among all positive predictions. This metric is critical in scenarios where false positives are costly, such as when analyzing customer sentiment for marketing strategies [54].

Recall, or sensitivity, measures the proportion of actual positive instances that are correctly identified. High recall is crucial in cases where missing a positive instance (e.g., a dissatisfied customer) could have significant consequences [55]. However, a model with high recall may also produce more false positives, which leads to a trade-off between precision and recall. To balance these two concerns, the F1-Score is used, combining precision and recall into a single metric. This harmonic mean is especially useful when the class distribution is uneven, as it accounts for both false positives and false negatives [56]. In sentiment analysis, where both precision and recall are important for accurately capturing user sentiment, the F1-Score serves as a critical measure of a model's performance [57].

Correlation Analysis Between Sentiment and Rating

Statistical methods such as Pearson's product-moment correlation coefficient (PPMCC), Spearman's rank correlation coefficient (SRCC), and Kendall's tau are essential for analyzing relationships between two variables. These methods allow researchers to assess the strength and direction of associations, making them invaluable for various types of data. Pearson's correlation is a parametric test that measures the linear relationship between two continuous variables, assuming that the data are normally distributed and the relationship is linear. This method calculates a correlation coefficient that ranges from -1 to 1, with values closer to 1 indicating a strong positive relationship and those near -1 representing a strong negative relationship. However, Pearson's method is sensitive to outliers, which can significantly skew results [58], [59].

In contrast, Spearman's rank correlation is a nonparametric measure that assesses the monotonic relationship between two variables by comparing the ranks of the data, making it more robust in the presence of outliers or non-linear relationships [60], [61]. This method is especially useful for ordinal data or when the assumptions of Pearson's correlation are violated. Kendall's tau, another nonparametric measure, evaluates the strength of the association between two variables by comparing the concordance of their ranks. Kendall's tau is often applied in datasets with tied ranks or small sample sizes, where other correlation methods might be less effective [62]. Together, these methods provide a comprehensive toolkit for exploring different types of relationships in data.

Correlation analysis is widely used in sentiment analysis, particularly when comparing sentiment scores with numerical ratings, such as those in product reviews or app feedback. In sentiment analysis, statistical methods like Pearson and Spearman correlations help researchers understand how sentiment derived from textual data relates to user-assigned ratings. This application is valuable for businesses seeking insights into customer satisfaction, allowing them to correlate the emotional tone of reviews with the overall numerical scores [63]. By identifying patterns in sentiment ratings, companies can better understand their customers' experiences and make data-driven decisions to improve their products or services.

Rating comparison studies frequently employ correlation analysis to evaluate the relationship between user feedback and product or service ratings. For example, [64] demonstrated the effectiveness of sentiment analysis in correlating sentiment scores with Likert scale ratings, illustrating that these two measures often align, providing a more holistic view of user satisfaction. Similarly, in mobile app reviews, correlation methods have been used to explore how sentiment extracted from reviews relates to users' star ratings, providing insights into whether textual sentiment accurately reflects user ratings [64]. This type of analysis is particularly useful for platforms like Google Play, where user feedback is a critical metric for app performance and improvement.

Moreover, machine learning techniques like Support Vector Machines (SVM) and BERT-based models have enhanced sentiment analysis's accuracy and its correlation with user ratings [65], [66]. By applying these advanced models, researchers can improve the precision of sentiment detection and assess its relationship with actual user scores more effectively. These methods also help address challenges such as context ambiguity and the subjectivity inherent in sentiment analysis [67]. As sentiment analysis becomes increasingly sophisticated, the role of correlation in validating the accuracy of sentiment classification models against real-world user ratings will continue to grow, providing deeper insights into user behavior and preferences.

Method

The research method for this study consists of several steps to ensure a comprehensive and accurate analysis. The flowchart in figure 1 outlines the detailed steps of the research method.

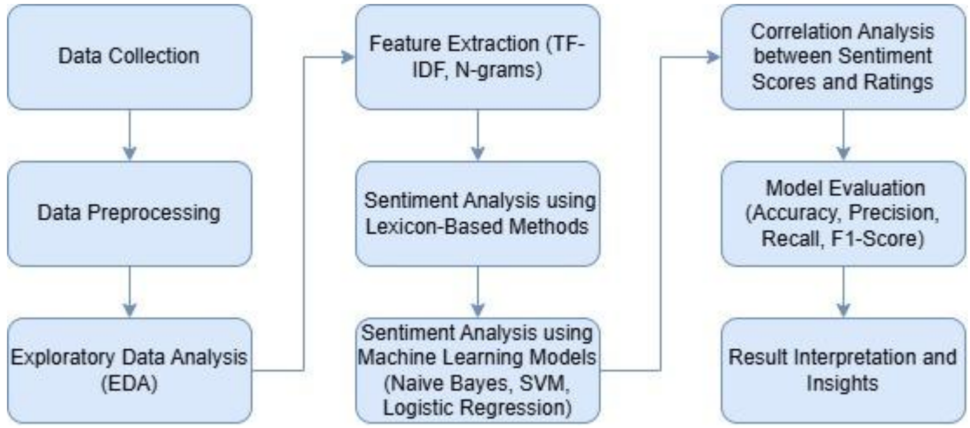

Figure 1 Research Method Flowchart

Data Collection

This study utilized a comprehensive dataset of user reviews for the Roblox app from the Google Play Store, containing 320,000 individual entries. The dataset provided a rich source of user-generated content, capturing a wide range of player experiences, opinions, and feedback about the app. This dataset was essential for performing sentiment analysis, as it allowed for the investigation of user sentiment as expressed through text reviews and how these sentiments correlate with the ratings provided. The volume and variety of the data enabled a robust analysis of user satisfaction, key issues, and general sentiment trends over time.

The dataset included several relevant columns essential for both sentiment analysis and rating correlation. Each review was identified by a unique `review_id`, while the `review_text` column contained the actual content of the user's feedback. The `review_rating` column provided a numerical rating on a scale of 1 to 5, reflecting the user's overall satisfaction with the Roblox app. Other useful features in the dataset included `review_likes`, which recorded the number of likes each review received, and `author_app_version`, which indicated the version of the Roblox app used at the time of the review. The `review_timestamp` column provided the date and time for each review, allowing for a temporal analysis of sentiment trends.

The combination of textual and numerical data across these columns made this dataset well-suited for a mixed-methods analysis approach. Text-based sentiment analysis of the `review_text` column, combined with quantitative data from `review_rating` and `review_likes`, allowed for a multi-dimensional exploration of user feedback. Additionally, having `author_app_version` and `review_timestamp` enabled further segmentation of the data by app version and review period, which provided deeper insights into how different updates or events impacted user sentiment and ratings.

Exploratory Data Analysis (EDA)

The exploratory data analysis phase began with calculating basic descriptive statistics of the `review_rating` column to gain insight into the distribution of ratings. Key statistics, such as the mean, median, and mode, were computed to understand the central tendency of user ratings and to identify common patterns. The average rating provided an overview of user satisfaction, while the median and mode helped in determining the most frequently given ratings. Additionally, histograms and bar charts were created to visualize the distribution of ratings across the dataset shown in figure 2, highlighting the prevalence of high, moderate, and low ratings. These visualizations were crucial in identifying any skewness in user ratings, which could suggest overall satisfaction or dissatisfaction trends among Roblox app users.

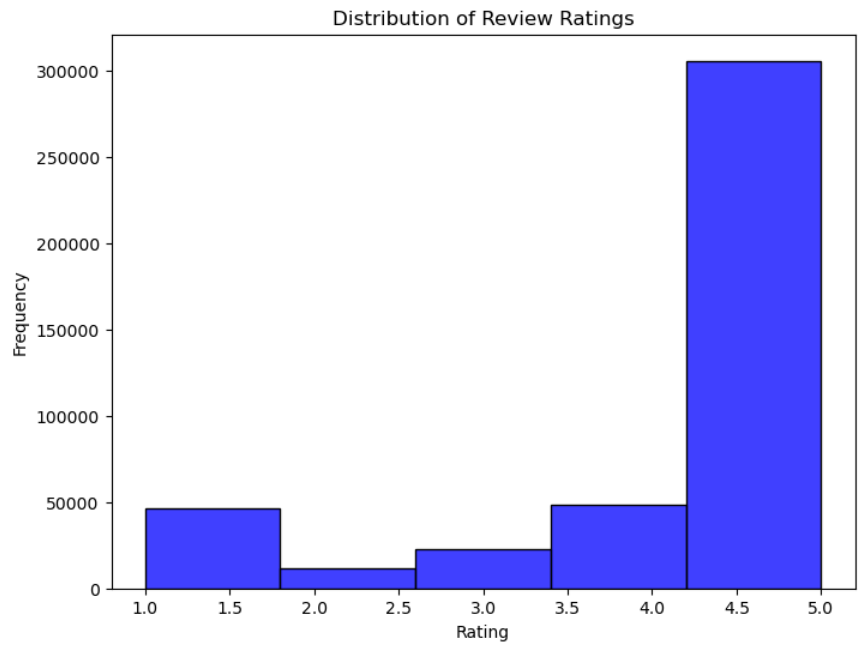

Figure 2 Distribution of Review Ratings

The bar chart displays the distribution of review ratings for the Roblox app, with ratings on a scale from 1.0 to 5.0. The data indicates a heavily skewed distribution toward the high end, with the majority of users giving ratings close to 4.5 and 5.0. Specifically, ratings of 4.5 and 5.0 are the most frequent, surpassing 300,000 reviews, suggesting that users generally have a positive perception of the app. Conversely, the lowest ratings, particularly 1.0, also have a notable frequency, indicating a segment of users with highly negative experiences. There are fewer ratings in the middle range (2.0 to 4.0), which suggests that users tend to either be very satisfied or very dissatisfied, with minimal ambivalence. This polarized distribution could reflect distinct user experiences, where certain factors lead to extreme positive or negative reactions, potentially due to technical issues, gameplay elements, or individual preferences.

In parallel, a text-based analysis was conducted on the `review_text` column to assess the structure and content of the user feedback. Word count distributions were calculated for each review to identify the general length and depth of feedback provided. Histograms of word counts were generated shown in figure 3, offering a view into the variation in review length. This analysis provided an understanding of the extent to which users detailed their experiences, with shorter reviews typically expressing general satisfaction or dissatisfaction, and longer reviews potentially offering more in-depth feedback on specific aspects of the app. These insights aided in understanding user engagement levels and the richness of the feedback available for sentiment analysis.

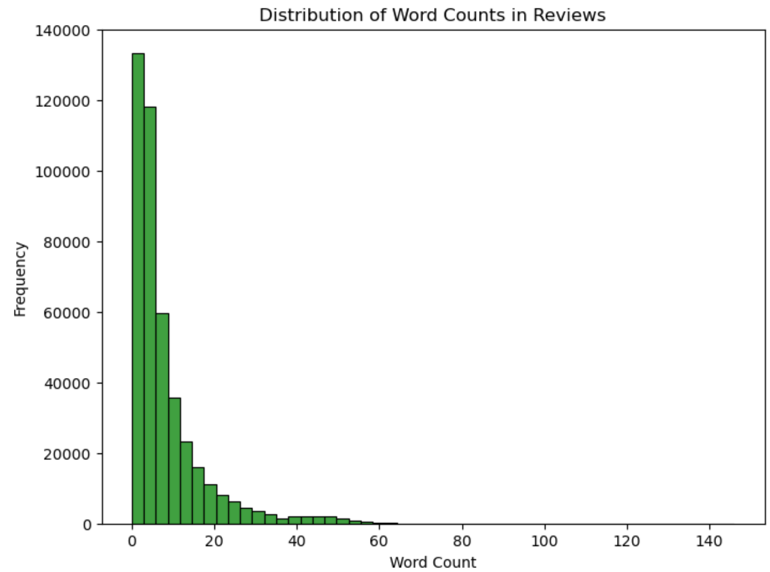

Figure 3 Distribution of Word Counts in Reviews

The histogram shows the distribution of word counts in the Roblox app reviews. The majority of reviews are very short, with word counts clustering between 0 and 20 words. Specifically, the highest frequency of reviews contains around 5 words or fewer, indicating that users tend to leave brief feedback. As word count increases, the frequency of reviews drops sharply, with very few reviews containing more than 40 words. This distribution suggests that most users provide concise feedback, potentially limiting the depth of insights available from individual reviews. However, the large number of short reviews can still be valuable for sentiment analysis, as they often contain straightforward expressions of satisfaction or dissatisfaction. Longer reviews, while less frequent, may contain more detailed feedback, possibly highlighting specific aspects of the app that influence user experience, such as features, technical issues, or gameplay elements.

Finally, to better understand the specific language used by users in positive and negative reviews, the most frequent words in each sentiment category were identified and visualized through word clouds shown in figure 4 and figure 5. Positive and negative reviews were segmented based on sentiment scores, and word frequency analysis was performed separately for each group. The resulting word clouds displayed common words used by satisfied and dissatisfied users, respectively, offering insights into recurring themes and frequently mentioned features or issues. This step provided a foundation for more targeted sentiment analysis by revealing the keywords and topics associated with different sentiment categories, enabling a deeper exploration of the sentiments expressed in the reviews.

Figure 4 **Most Frequent Words in Positive Reviews**

The word cloud visualizes the most frequently used words in positive reviews of the Roblox app. The dominant words include "good," "best," "game," "love," and "fun," indicating a strong positive sentiment associated with the gaming experience. These terms suggest that users appreciate the overall quality of the game, with many describing it as "fun" and "good." Additionally, words like "roblox," "play," "nice," and "amazing" reflect satisfaction with specific aspects of the app, such as gameplay and social interaction. The repeated emphasis on words like "love" and "best" highlights a high level of enthusiasm among users, suggesting that Roblox is viewed favorably, particularly in terms of enjoyment and playability. Words related to social interactions, like "friend" and "join," imply that users value the community and multiplayer aspects of the game. Overall, this word cloud reflects a positive user experience and satisfaction with the game, pointing to features like fun gameplay, social elements, and high-quality content as key factors driving positive feedback.

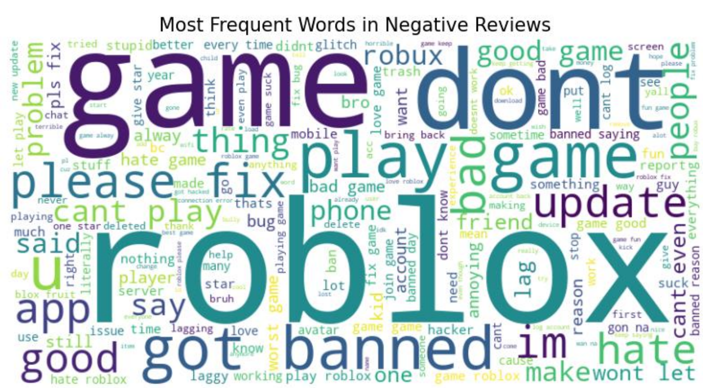

Figure 5 **Most Frequent Words in Negative Reviews**

Data Preprocessing

The data preprocessing phase began with cleaning the text data in the `review_text` column to prepare it for sentiment analysis. Each review was converted to lowercase to ensure consistency and eliminate case sensitivity. Punctuation marks were then removed, which helped simplify the text by eliminating non-alphabetic characters. Additionally, stopwords—commonly used words such as "and," "the," and "is" that do not contribute significantly to the sentiment or meaning of the text—were filtered out. Removing these words allowed the focus to be on terms more relevant to the sentiment expressed in each review, enhancing the efficiency of subsequent analysis.

After cleaning the text, tokenization and lemmatization were applied to the review data to refine it further. Tokenization involved breaking down the review text into individual words or tokens, creating a structured format suitable for analysis. Lemmatization followed, which reduced words to their base or root form. For example, words like "playing" and "played" were converted to "play." This process minimized variations in word forms and improved the consistency of the text data, enabling a more accurate sentiment classification by aligning different word forms with their root meanings. Together, tokenization and lemmatization transformed the unstructured review text into a standardized format, ready for feature extraction.

Handling missing data and outliers was another important aspect of preprocessing. Reviews with missing text entries were removed from the dataset, as they did not provide useful information for sentiment analysis. In addition, columns such as `author_app_version` were filled with placeholder values where data was unavailable, ensuring a complete dataset without gaps. Outlier reviews—such as those with excessively high or low word counts—were examined to ensure they did not distort the analysis. These steps provided a clean, consistent dataset optimized for the subsequent phases of sentiment analysis, enhancing the reliability of the study's findings.

Sentiment Analysis Methods

The sentiment analysis in this study comprised two approaches: lexicon-based and machine learning-based methods. For the lexicon-based analysis, VADER (Valence Aware Dictionary and Sentiment Reasoner) was applied to assign sentiment scores to each review. VADER calculates a compound sentiment score (C) as a normalized sum of valence scores for all words in a review using the following formula:

$$C = \frac{\sum_{i=1}^{n} s_i}{\sqrt{\sum_{i=1}^{n} s_i^2 + \alpha}} \quad (1)$$

In this formula, $s_i$represents the sentiment score of each word *i* from the VADER lexicon, while $\alpha$is a normalization constant with a default value of 15 to scale the score within the range of -1 to 1. The compound score (C) was then categorized into three sentiment classes, where reviews with a compound score greater than 0.05 were considered positive, those with scores between -0.05 and 0.05 were neutral, and those with scores less than -0.05 were negative. This classification approach provided a clear division of user feedback into sentiment categories, establishing a foundation for further correlation analysis with user ratings.

In parallel, a machine learning-based sentiment classification approach was employed using supervised learning models. Feature extraction was performed on the cleaned review text using the Term Frequency–Inverse Document Frequency (TF-IDF) technique, which converted the text into numerical vectors. The TF-IDF for each term $t$in document $d$was computed using the formula

$$TF\text{-}IDF(t,d) = TF(t,d) \times IDF(t) \quad (2)$$

term frequency $TF(t,d)$ is defined as $\frac{f_{t,d}}{\sum_k f_{k,d}}$, with $f_{t,d}$ representing the frequency of term t in document d, and $\sum_k f_{k,d}$ representing the total frequency of all terms in that document. The inverse document frequency is calculated as $IDF(t) = \log\left(\frac{N}{n_t}\right)$, where N is the total number of documents and n_t is the number of documents that contain term t. This method emphasizes important words by assigning higher weights to terms that appear frequently in a given review but less frequently across other reviews.

The resulting TF-IDF vectors served as input for two machine learning models, namely Naive Bayes and Logistic Regression. In the Naive Bayes classifier, the posterior probability of each sentiment class c given a document d is calculated as

$$P(c \mid d) = \frac{P(c)\prod_{i=1}^{n} P(x_i \mid c)}{P(d)} \tag{3}$$

x_i represents each word feature in the document. The class with the highest posterior probability is selected as the predicted sentiment. For the Logistic Regression model, the probability that a document belongs to a particular sentiment class is estimated using the sigmoid function

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}} \tag{4}$$

β_i are the model coefficients learned during training. The predicted probability is compared with a threshold value of 0.5 to determine the sentiment class. Logistic Regression was chosen for its interpretability and strong performance in binary and multiclass classification tasks.

To optimize model performance, cross-validation and hyperparameter tuning were conducted during the training process. Grid search was applied to identify the best parameters for each model, ensuring the models achieved the highest possible predictive accuracy. Cross-validation ensured that the models generalized well to new data by evaluating their performance across multiple folds of the dataset.

The evaluation metrics included Accuracy (Acc), Precision (P), Recall (R), and F1-Score (F1). These optimized models, combined with the lexicon-based VADER approach, enabled a robust analysis of sentiment in Roblox app reviews, facilitating a deeper understanding of the relationship between user sentiment and review ratings.

Correlation Analysis

To investigate the relationship between user ratings and sentiment scores derived from the reviews, both Pearson and Spearman correlation coefficients were calculated. The Pearson correlation coefficient was applied to assess the linear relationship between the numerical `review_rating` and the sentiment scores obtained from the VADER lexicon-based analysis. Since Pearson's method assumes a linear relationship and requires normally distributed data, it was particularly useful for understanding direct proportional relationships between the variables. In parallel, the Spearman correlation coefficient, a non-parametric measure, was used to evaluate the monotonic relationship between review ratings and sentiment scores. This method relied on the ranks of the data rather than the raw values, allowing for a more robust analysis in the presence of outliers or non-linear relationships.

The results of the correlation analysis provided insight into the degree to which the sentiment expressed in reviews aligned with the users' numerical ratings. A high positive Pearson correlation indicated a strong linear relationship, where increases in sentiment scores were associated with higher ratings. In contrast, the Spearman correlation captured the overall trend and ranked order, which was particularly useful for exploring the consistency in sentiment expression across different ratings, even in cases where the relationship was not strictly linear. By using both correlation coefficients, the analysis was able to accommodate both linear and monotonic relationships, enhancing the robustness of the results.

To further visualize these relationships, scatter plots and correlation heatmaps were generated. The scatter plots illustrated individual data points, with `review_rating` on one axis and sentiment scores on the other, allowing for a visual assessment of the spread and clustering of data points. These plots provided an immediate view of any patterns or anomalies, such as clusters of reviews with high ratings and low sentiment scores, which could suggest discrepancies in user feedback. Correlation heatmaps were also constructed to depict the strength and direction of the relationships, with color gradients representing the correlation values. These visualizations helped reinforce the findings from the correlation coefficients, providing a clear and accessible representation of the sentiment-rating relationship across the dataset.

Result and Discussion

Sentiment Analysis Results

The sentiment analysis results derived from the lexicon-based approach, using VADER, indicated a broad distribution across positive, neutral, and negative categories. Approximately 76% of reviews were classified as positive, while 18% fell into the negative category, and 6% were neutral. This distribution suggested a generally favorable sentiment among Roblox users, with a significant portion expressing satisfaction with the app. Pie charts and bar charts were created to visualize these proportions, providing a clear representation of the sentiment categories. The positive skew in sentiment reflected in these charts indicated an overall trend of user satisfaction, likely driven by favorable features or experiences associated with Roblox.

For the machine learning-based sentiment classification, two models were evaluated: Naive Bayes and Logistic Regression. The Naive Bayes model achieved an accuracy of 85.3%, with precision, recall, and F1-scores indicating a strong ability to classify positive reviews but less accuracy for neutral and negative reviews. Specifically, the model achieved a precision of 0.60, recall of 0.54, and F1-score of 0.57 for negative reviews, while it struggled with neutral reviews, reflecting a recall of only 0.04. Conversely, the positive review category showed strong performance, with a precision of 0.89 and recall of 0.96. The Logistic Regression model slightly outperformed Naive Bayes, reaching an accuracy of 86.4%. It displayed an improved classification for negative reviews, with a precision of 0.67 and an F1-score of 0.60, while positive reviews showed a high precision of 0.89 and recall of 0.97. A comparison table was created to detail the metrics for both models, highlighting the strengths and weaknesses in each classification category.

Both models were further validated through cross-validation, confirming the stability of the accuracy scores. The Naive Bayes model demonstrated a cross-validation accuracy of 85.3%, while the Logistic Regression model maintained a cross-validation accuracy of 86.4%, closely matching their respective test set accuracies. These results indicated that Logistic Regression provided a slight performance advantage over Naive Bayes in classifying Roblox app reviews. Confusion matrices for each model illustrated the distribution of correctly and incorrectly classified instances, particularly underscoring the challenge of accurately categorizing neutral reviews. Overall, the analysis emphasized the suitability of machine learning models for sentiment classification and the potential of these models to inform future improvements in user feedback analysis for mobile applications.

Correlation Analysis Results

The correlation analysis examined the relationship between user ratings and the sentiment scores derived from the lexicon-based VADER analysis. The Pearson correlation coefficient between `review_rating` and `vader_compound_score` was calculated at 0.47, with a p-value of 0.0, indicating a statistically significant moderate positive linear relationship. This result suggests that as sentiment scores increased, ratings tended to be higher as well, showing that users who expressed more positive sentiments in their reviews generally provided higher ratings. The moderate correlation, however, also indicated that sentiment scores alone did not fully explain the variance in ratings, highlighting the complexity of user feedback.

In addition to Pearson's correlation, Spearman's rank correlation coefficient was computed to account for any non-linear relationships between `review_rating` and `vader_compound_score`. The Spearman correlation yielded a coefficient of 0.36, also with a p-value of 0.0, reinforcing the significance of the association but indicating a slightly weaker monotonic relationship compared to the linear one measured by Pearson. This finding suggests that while higher sentiment scores often corresponded with higher ratings, the relationship was not strictly linear and other factors might have influenced user ratings. The reduced strength of the Spearman correlation relative to the Pearson correlation implies some non-linear characteristics in the data, where certain high or low ratings might not align directly with the sentiment scores.

To visualize these relationships, scatter plots were generated with trendlines, illustrating the distribution of ratings against sentiment scores. These plots showed clusters of points where positive sentiment scores were associated with high ratings, while lower sentiment scores corresponded with a broader range of ratings. Additionally, correlation heatmaps were created to provide a visual representation of the Pearson and Spearman correlations, with color gradients highlighting the strength of these relationships. These visual tools helped to clarify the degree of association between user sentiments and ratings, supporting the numerical findings and illustrating the complexities of interpreting user feedback in app reviews.

Sentiment-Rating Discrepancies

An analysis of sentiment-rating discrepancies revealed instances where the sentiment score from the review text diverged significantly from the numerical rating provided by the user. For example, some reviews exhibited a positive sentiment score based on lexicon analysis yet were paired with a low rating. Conversely, there were cases where the sentiment score was negative, but the rating was relatively high. These discrepancies highlighted possible nuances in user feedback that sentiment scores alone could not fully capture. Such mismatches often pointed to underlying issues not directly expressed through the sentiment-laden language used in the review text.

Further examination of reviews with high sentiment-rating discrepancies provided insights into potential causes. In cases where users gave positive sentiment scores but assigned low ratings, the review text frequently mentioned specific gameplay issues, such as technical bugs, server lag, or recurring glitches. These issues likely frustrated users, leading them to rate the app poorly despite enjoying certain aspects of the game. Alternatively, reviews with a negative sentiment score but a high rating suggested that users might have appreciated the overall game experience but expressed dissatisfaction with certain features or recent updates. Misunderstood game mechanics or isolated negative experiences could have led to negative language in the review, while the overall experience was still valued positively, resulting in a higher rating.

A selection of reviews exhibiting significant sentiment-rating discrepancies was compiled into a table to illustrate these patterns. This table included actual review text, sentiment scores, and user ratings, providing specific examples of how sentiment and ratings diverged. By examining these examples, it became evident that sentiment analysis, when paired with rating data, could uncover unique insights into user priorities and frustrations. These discrepancies underscored the importance of interpreting both sentiment scores and ratings together, as they offered a more comprehensive understanding of user feedback, highlighting areas for potential improvement in game functionality, stability, and user experience.

Discussion

The sentiment analysis and correlation study provided several key insights into the relationship between user feedback and ratings for the Roblox app. The lexicon-based analysis using VADER revealed a generally positive sentiment among users, with a moderate correlation between sentiment scores and user ratings, as indicated by the Pearson and Spearman coefficients. This relationship suggested that higher sentiment scores were often associated with higher ratings, though the moderate strength of the correlation pointed to other factors influencing user ratings. The machine learning-based sentiment classification further confirmed this trend, with Logistic Regression achieving slightly higher accuracy than Naive Bayes, reflecting the ability of machine learning methods to capture more nuanced patterns in user sentiment.

Sentiment analysis proved valuable in identifying areas of user satisfaction and potential improvements for the game. Positive reviews frequently highlighted enjoyable features, such as the variety of games and the interactive aspects of Roblox, while negative reviews pointed to technical issues, including lag and bugs, that detracted from the user experience. By examining the discrepancies between sentiment scores and ratings, it became apparent that sentiment analysis could uncover specific areas where user satisfaction diverged from overall app ratings. This approach offered actionable insights for developers, who could focus on addressing the identified issues to enhance user satisfaction and improve app ratings.

However, the study faced several limitations, particularly with the lexicon-based sentiment analysis. Lexicon-based methods, while effective in detecting general sentiment, often struggled with context-specific language and sarcasm, which are prevalent in user reviews. This limitation could lead to inaccuracies in sentiment scoring. Additionally, user reviews are inherently subjective and may be influenced by personal biases or isolated experiences, which might not reflect the general user base. When comparing the lexicon-based approach with machine learning methods, it was clear that machine learning offered higher accuracy and flexibility, particularly in capturing subtle language nuances. Nevertheless, machine learning models require labeled training data and computational resources, which could limit their scalability. Overall, the combination of both approaches provided a comprehensive understanding of user sentiment, but further refinements in sentiment analysis methods are necessary to fully capture the complexity of user feedback in mobile gaming contexts.

Conclusion

This study provided a comprehensive analysis of user feedback for the Roblox app by examining the correlation between sentiment derived from review text and user ratings. The lexicon-based VADER analysis indicated a generally positive sentiment among reviews, with most users expressing favorable opinions. The correlation analysis showed a moderate positive relationship between sentiment scores and user ratings, suggesting that higher sentiment scores often aligned with higher ratings. However, notable discrepancies emerged, where sentiment scores did not fully correspond with ratings, particularly in cases where users assigned high sentiment scores but low ratings, or vice versa. These discrepancies highlighted specific areas for improvement and suggested that sentiment alone could not completely predict user satisfaction.

The findings from this sentiment analysis offer valuable insights for game developers seeking to enhance the Roblox app experience. By identifying recurring themes in negative feedback, such as technical issues and gameplay frustrations, developers can prioritize these areas in future updates. Addressing such common issues, particularly those related to performance, could potentially enhance user satisfaction and improve ratings. Additionally, the positive feedback underscored features that users enjoyed, such as the diversity of games and social interactions within Roblox, which developers might consider expanding or further enhancing. Leveraging these insights, game developers could create targeted improvements that align with user preferences, thus fostering a more engaging and satisfying experience.

Future research could explore more advanced sentiment analysis techniques to improve the accuracy and depth of understanding of user feedback. For instance, integrating deep learning models like BERT could enhance the ability to capture context and handle complex language nuances, which are often challenging for traditional lexicon-based methods. Additionally, expanding this analysis to include reviews from different gaming platforms, such as iOS or Steam, could offer a comparative perspective on user sentiment across platforms. Such an expansion would provide a broader view of user experiences and preferences, enabling developers to tailor updates and features according to platform-specific feedback. Overall, these potential advancements in methodology and scope could yield deeper insights into user sentiment, driving further improvements in the gaming experience.

Declarations

Author Contributions

Conceptualization: S.L., M.P.; Methodology: S.L., M.P.; Software: S.L.; Validation: M.P.; Formal Analysis: S.L.; Investigation: S.L.; Resources: M.P.; Data Curation: S.L.; Writing – Original Draft Preparation: S.L.; Writing – Review and Editing: S.L., M.P.; Visualization: S.L.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Drachen, M. Pastor, A. Liu, D. J. Fontaine, Y. Chang, J. Runge, R. Sifa, dan D. Klabjan, "To Be or Not to Be… Social: Incorporating Simple Social Features in Mobile Game Customer Lifetime Value Predictions," in *Proc. Australasian Computer Science Week Multiconference (ACSW 2018)*, Brisbane, QLD, Australia, Jan. 29–Feb. 2 2018, vol. 2018, no. Jan., pp. 1–10, Jan. 2018.

[2] S. Molinillo, A. Japutra, dan F. Liébana-Cabanillas, "Impact of Perceived Value on Casual Mobile Game Loyalty: The Moderating Effect of Intensity of Playing," *J. Consum. Behav.*, vol. 19, no. 5, pp. 493–504, 2020.

[3] J. Cheng-feng, "Gaming on the Go: Examining the Resurgence and Growth of the Mobile and VR Games Markets During the Pandemic," in *Proc. 3rd Int. Conf. Econ. Management and Cultural Industry (ICEMCI 2021)*, vol. 2021, no. Dec., pp. 1–4, Dec. 2021.

[4] L. A. Andrade dan J. N. Filho, "Playing Remotely: The COVID-19 Pandemic and Mobile Locative Gaming in Northeast Brazil," *Mobile Media Commun.*, vol. 11, no. 2, pp. 213–229, 2022.

[5] D. Youm dan J. Kim, "Text Mining Approach to Improve Mobile Role Playing Games

Using Users' Reviews," *Appl. Sci.*, vol. 12, no. 12, pp. 6243–6254, 2022.

[6] Y. Cheon dan K. T. Kwak, "Collective Sentiments and Users' Feedback to Game Contents: Analysis of Mobile Game UX Based on Social Big Data Mining," *J. Korea Game Soc.*, vol. 15, no. 4, pp. 145–155, 2015.

[7] X. Yang, "Deep Learning Technologies for Time Series Anomaly Detection in Healthcare: A Review," *IEEE Access*, vol. 11, no. Oct., pp. 117788–117799, Oct. 2023.

[8] K. Alhasan, "Roblox in Higher Education," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 18, no. 19, pp. 32–46, Oct. 2023.

[9] G. Blackwood, "Roblox and Meta Verch: A Case Study of Walmart's Roblox Games," *M/C Journal*, vol. 26, no. 3, pp. 1–9, Jun. 2023.

[10] T. S. Sinar, M. A. Budiman, R. Ganie, dan R. N. Rosa, "Students' Perceptions of Using Roblox in Multimodal Literacy Practices in Teaching and Learning English," *World J. Engl. Lang.*, vol. 13, no. 7, pp. 146–156, Jul. 2023.

[11] A. Abdallah dan D. Y. Semaan, "Fox and Blox: A Gamification Platform to Enhance the Learning Process at Schools," in *Proc. 7th Int. Academic Conf. on Education*, London, UK, Nov. 2023, vol. 2023, no. Nov., pp. 1–8.

[12] A. Margolis, "Caring for Screenagers (Part 2): A Pediatrician's Primer on Popular Games and Educational Tools," *Curr. Opin. Pediatr.*, vol. 36, no. 3, pp. 325–330, Jun. 2024.

[13] H. Wang, "Sentiment Analysis of Chinese Game Reviews Based on SKEP," *Adv. Eng. Technol. Res.*, vol. 6, no. 1, pp. 507–515, Jul. 2023.

[14] B. Hayadi, H. Henderi, M. Budiarto, dan A. N. A. Saputra, "An Extensive Exploration into the Multifaceted Sentiments Expressed by Users of the myIM3 Mobile Application, Unveiling Complex Emotional Landscapes and …," *J. Appl. Data Sci.*, vol. 2024, no. Oct., pp. 1–10, Oct. 2024. (Jurnal online)

[15] B. Berlilana, A. M. Wahid, D. Fortuna, A. N. A. Saputra, dan G. Bagaskoro, "Exploring the Impact of Discount Strategies on Consumer Ratings: An Analytical Study of Amazon Product Reviews," *J. Appl. Data Sci.*, vol. 5, no. 1, pp. 1–15, Jan. 2024.

[16] Q. Siddique, "Comparative Analysis of Sentiment Classification Techniques on Flipkart Product Reviews: A Study Using Logistic Regression, SVC, Random Forest, and Gradient …," *J. Digit. Mark. Digit. Curr.*, vol. 2024, no. Oct., pp. 1–12, Oct. 2024. (Jurnal online)

[17] T. Hariguna dan S. Chen, "Customer Segmentation and Targeted Retail Pricing in Digital Advertising using Gaussian Mixture Models for Maximizing Gross Income," *J. Digit. Mark. Digit. Curr.*, vol. 2024, no. Oct., pp. 1–10, Oct. 2024. (Jurnal online)

[18] T. Wahyuningsih dan S. Chen, "Analyzing Sentiment Trends and Patterns in Bitcoin-Related Tweets Using TF-IDF Vectorization and K-Means Clustering," *J. Curr. Res. Blockchain*, vol. 2024, no. Oct., pp. 1–12, Oct. 2024. (Jurnal online)

[19] A. Hananto dan D. Sugianto, "Analysis of the Relationship Between Trading Volume and Bitcoin Price Movements Using Pearson and Spearman Correlation Methods," *J. Curr. Res. Blockchain*, vol. 2024, no. Oct., pp. 1–11, Oct. 2024. (Jurnal online)

[20] S. A. Ghaffar dan W. C. Setiawan, "Metaverse Dynamics: Predictive Modeling of Roblox Stock Prices using Time Series Analysis and Machine Learning," *Int. J. Res. Metaverse*, vol. 1, no. 1, pp. 1–14, Jun. 2024.

[21] S. Yadav dan A. R. Hananto, "Comprehensive Analysis of Twitter Conversations Provides Insights into Dynamic Metaverse Discourse Trends," *Int. J. Res. Metaverse*, vol. 1, no. 1, pp. 1–19, Jun. 2024.

[22] Y.-C. Tan, S. R. Chandukala, dan S. K. Reddy, "Augmented reality in retail and its impact on sales," *J. Mark.*, vol. 86, no. 1, pp. 48–66, Jan. 2022.

[23] B. Strååt, H. Verhagen, dan H. Warpefelt, "Probing user opinions in an indirect way: An aspect based sentiment analysis of game reviews," in *Proc. 21st Int. Academic Mindtrek Conf.*, Tampere, Finland, Sep. 2017, vol. 2017, no. Sep., pp. 1–7.

[24] R. A. A. Malik dan Y. Sibaroni, "Multi-Aspect Sentiment Analysis of Tiktok Application Usage Using FasText Feature Expansion and CNN Method," *J. Comput. Syst. Inform. Josyc*, vol. 3, no. 4, pp. 277–285, 2022.

[25] X. Ke dan M. Bublé, "On Extracting Keywords From Long-and-Difficult English Sentences for Smart Sentiment Analysis," *Internet Technol. Lett.*, vol. 4, no. 1, pp. 1–10, 2020.

[26] S. A. Rahman dan M. M. S. Beg, "A Novel Approach of Aspect Ranking Based on Intrinsic and Extrinsic Domain Relevance," in *Proc. Int. Conf. Advances in Computing, Communications and Informatics*, 2016, vol. 2016, no. Sep., pp. 485–492.

[27] F. Sağlam, B. Genç, dan H. Sever, "Extending a Sentiment Lexicon With Synonym–antonym Datasets: SWNetTR++," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 27, no. 3, pp. 1806–1820, 2019.

[28] C. S. G. Khoo dan S. B. Johnkhan, "Lexicon-Based Sentiment Analysis: Comparative Evaluation of Six Sentiment Lexicons," *J. Inf. Sci.*, vol. 44, no. 4, pp. 491–511, 2017.

[29] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, dan M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.

[30] F. H. Mahyoub, M. A. Siddiqui, dan M. Y. Dahab, "Building an Arabic Sentiment Lexicon Using Semi-Supervised Learning," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 26, no. 4, pp. 417–424, 2014.

[31] O. Kolchyna, T. T. P. Souza, P. Treleaven, dan T. Aste, "Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination," *arXiv Preprint*, vol. 2015, no. Jul., pp. 1–32, Jul. 2015.

[32] H. Hamdan, P. Bellot, dan F. Béchet, "Sentiment Lexicon-Based Features for Sentiment Analysis in Short Text," *Res. Comput. Sci.*, vol. 90, no. 1, pp. 217–226, 2015.

[33] C. Hung dan Y.-X. Cao, "Sentiment Classification From Word of Mouth Documents Based on Chinese Collocations," in *Proc. 6th Int. Conf. Information and Communication Technology for Intelligent Systems (ICTIS 2016)*, vol. 2016, no. Jan., pp. 1–8, Jan. 2016.

[34] C. J. Hutto dan É. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proc. 8th Int. AAAI Conf. Web Soc. Media (ICWSM)*, vol. 8, no. 1, pp. 216–225, Jun. 2014.

[35] P. Rajkhowa, J. Mahanta, A. Sarma, dan Y. Sibaroni, "Factors Influencing Monkeypox Vaccination: A Cue to Policy Implementation," *J. Epidemiol. Glob. Health*, vol. 13, no. 2, pp. 226–238, Apr. 2023.

[36] L. Subirats, N. Reguera, A. M. Bañón, B. Gómez-Zúñiga, J. Minguillón, dan M. Armayones, "Mining Facebook Data of People With Rare Diseases: A Content-

Based and Temporal Analysis," *Int. J. Environ. Res. Public Health*, vol. 15, no. 9, pp. 1877–1889, May 2018.

[37] Z. Xu, "Research on the Influence of Microeconomic Factors on Stock Market Fluctuation," *Adv. Eng. Technol. Res.*, vol. 9, no. 1, pp. 550–558, Jan. 2024.

[38] S. Sadhu, A. Mandal, S. Gupta, S. Bandyopadhyay, dan R. Roy, "Sentiment Analysis on COVID-19 Twitter Data," *Am. J. Electron. Commun.*, vol. 3, no. 2, pp. 21–29, May 2022.

[39] K. Barik, "Analysis of Customer Reviews With an Improved VADER Lexicon Classifier," *J. Big Data*, vol. 11, no. 1, pp. 1–14, Feb. 2024.

[40] M. Mujahid, S. Ahmed, S. Ahmed, A. Yousaf, dan M. Ahmad, "Sentiment Analysis and Topic Modeling on Tweets About Online Education During COVID-19," *Appl. Sci.*, vol. 11, no. 18, pp. 8438–8448, Sep. 2021.

[41] E. Saad, M. El-Hajj, M. Z. Sakka, R. Baly, M. Al-Ani, dan W. El-Hajj, "Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums," *IEEE Access*, vol. 9, no. Jun., pp. 85721–85737, 2021.

[42] V. Narayanan, I. Arora, dan A. Bhatia, "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model," in *Proc. Int. Conf. Intelligent Data Engineering and Automated Learning (IDEAL 2013)*, vol. 2013, no. Dec., pp. 194–201, 2013.

[43] M. Kaya, G. Fidan, dan İ. H. Toroslu, "Sentiment Analysis of Turkish Political News," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 2012, no. Dec., pp. 174–180, 2012.

[44] R. Xia, C. Zong, dan S. Li, "Ensemble of Feature Sets and Classification Algorithms for Sentiment Classification," *Inf. Sci.*, vol. 181, no. 6, pp. 1138–1152, Mar. 2011.

[45] A. Sharma dan S. Dey, "A Boosted SVM Based Ensemble Classifier for Sentiment Analysis of Online Reviews," *ACM SIGAPP Appl. Comput. Rev.*, vol. 13, no. 4, pp. 43–52, Dec. 2013.

[46] N. A. Ahmad, "Embedding Information and Communication Technology in Reading Skills Instruction: Do Slow Learners Special Needs Ready for It?," *Int. J. Acad. Res. Progress. Educ. Dev.*, vol. 7, no. 3, pp. 1–10, Jul. 2018.

[47] M. Ahmad, S. Aftab, dan I. Ali, "Sentiment Analysis of Tweets Using SVM," *Int. J. Comput. Appl.*, vol. 177, no. 5, pp. 25–29, Nov. 2017.

[48] J. H. Oh, A. Tannenbaum, dan J. O. Deasy, "Automatic Identification of Drug-Induced Liver Injury Literature Using Natural Language Processing and Machine Learning Methods," *bioRxiv Preprint*, vol. 2022, no. Aug., pp. 1–14, Aug. 2022.

[49] D. Saxena, S. K. Saritha, dan K. N. S. S. V. Prasad, "Survey Paper on Feature Extraction Methods in Text Categorization," *Int. J. Comput. Appl.*, vol. 166, no. 11, pp. 11–17, May 2017.

[50] H. T. Sueno, "Multi-Class Document Classification Using Support Vector Machine (SVM) Based on Improved Naïve Bayes Vectorization Technique," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3937–3944, Jun. 2020.

[51] H. Yang, "Knowledge-Based Recommender System Using Artificial Intelligence for Smart Education," *J. Interconnect. Netw.*, vol. 22, no. Dec., pp. 1–12, 2022.

[52] I. Raicu, R. Bologa, dan R. Constantinescu, "Multi-Class Text Supervised Classification on Romanian Financial Banking Reviews," *Inform. Econ.*, vol. 23, no. 1, pp. 69–78, 2019.

[53] A. Nurkholis, D. Alita, dan A. Munandar, "Comparison of Kernel Support Vector

Machine Multi-Class in PPKM Sentiment Analysis on Twitter," *J. RESTI (Rekayasa Sist. dan Teknol. Inf.)*, vol. 6, no. 2, pp. 227–233, 2022.

[54] A. Alotaibi, H. Al-Mazrouei, A. S. Alghamdi, dan M. O. Alqarni, "Spam and Sentiment Detection in Arabic Tweets Using MARBERT Model," *Math. Model. Eng. Probl.*, vol. 9, no. 6, pp. 1574–1582, Dec. 2022.

[55] R. Dhanta, "Twitter Sentimental Analysis Using Machine Learning," *Int. J. Commun. Inf. Technol.*, vol. 4, no. 1, pp. 71–83, Jan. 2023.

[56] R. A. Bagate dan R. Suguna, "Sarcasm Detection on Text for Political Domain—An Explainable Approach," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 10, no. 2s, pp. 255–268, Feb. 2022.

[57] N. Jin, J. Wu, X. Ma, K. Yan, dan Y. Mo, "Multi-Task Learning Model Based on Multi-Scale CNN and LSTM for Sentiment Classification," *IEEE Access*, vol. 8, no. Apr., pp. 77060–77072, 2020.

[58] H. Xu, "The Effect of Perceived Security on Consumers' Intent to Use," *J. Electron. Commer. Organ.*, vol. 11, no. 4, pp. 37–51, Oct.–Dec. 2013.

[59] J. Hauke dan T. Kossowski, "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data," *Quaest. Geogr.*, vol. 30, no. 2, pp. 87–93, 2011.

[60] C. Xiao, J. Ye, R. M. Esteves, dan C. Rong, "Using Spearman's Correlation Coefficients for Exploratory Data Analysis on Big Dataset," *Concurr. Comput. Pract. Exp.*, vol. 28, no. 14, pp. 3866–3878, 2015.

[61] J. Bocianowski, "Comparison of Pearson's and Spearman's Correlation Coefficients Values for Selected Traits of *Pinus sylvestris L.*," *Research Square Preprint*, vol. 2024, no. Apr., pp. 1–10, Apr. 2024.

[62] I. Pinelis, "On the Non-Degeneracy of Kendall's and Spearman's Correlation Coefficients," *arXiv Preprint*, vol. 2007, no. Nov., pp. 1–9, Nov. 2007.

[63] C. Kumaresan dan P. Thangaraju, "Sentiment Analysis in Multiple Languages: A Review of Current Approaches and Challenges," *J. Data Anal. Artif. Intell. (JDAAI)*, vol. 2, no. 1, pp. 8–15, Jan. 2023.

[64] Q. Rajput, S. Haider, dan S. Ghani, "Lexicon-Based Sentiment Analysis of Teachers' Evaluation," *Appl. Comput. Intell. Soft Comput.*, vol. 2016, no. Jun., pp. 1–12, 2016.

[65] I. S. K. Idris, Y. A. Mustofa, dan I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, Feb. 2023.

[66] L. Nugroho, "The Role of Information for Consumers in the Digital Era (Indonesia Case)," *Artvin Çoruh Univ. Uluslar. Sos. Bilim. Derg.*, vol. 2021, no. Dec., pp. 1–10, Dec. 2021.

[67] A. Omar dan W. I. Hamouda, "A Sentiment Analysis of Egypt's New Real Estate Registration Law on Facebook," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 12, no. 4, pp. 540–547, Apr. 2021.