ARAMBEPOLA, N., MUNASINGHE, L. and WARNAJITH, N. 2024. Factors influencing mobile app user experience: an analysis of education app user reviews. In *4th International conference on advanced research in computing 2024 (ICARC 2024), 21-24 February 2024, Belihuloya, Sri Lanka*. Piscataway: IEEE [online], pages 223-228. Available from: https://doi.org/10.1109/ICARC61713.2024.10499727

Factors influencing mobile app user experience: an analysis of education app user reviews.

ARAMBEPOLA, N., MUNASINGHE, L. and WARNAJITH, N.

2024

Factors Influencing Mobile App User Experience: An Analysis of Education App User Reviews

Nimasha Arambepola
Faculty of Science
University of Kelaniya
Kelaniya, Sri Lanka
nimasha@kln.ac.lk

Lankeshwara Munasinghe
School of Computing
Robert Gordon University
Aberdeen, Scotland, United Kingdom
l.munasinghe@rgu.ac.uk

Nalin Warnajith
Faculty of Science
University of Kelaniya
Kelaniya, Sri Lanka
nwarnajith@kln.ac.lk

***Abstract*—In the competitive digital world, user reviews considered as the most vital source of user feedback, provide valuable insights that reflect the success of software applications in terms of user experience (UX). As user-generated content grows exponentially, extracting meaningful information from user reviews has become an immensely challenging task. Though existing approaches can identify UX factors from mobile app reviews with a certain accuracy, prioritizing these factors poses a significant challenge. This research proposes a method to identify influential UX factors for mobile app reviews. Specifically, we did an in-depth analysis on educational app reviews of the Google Play Store. Notably, it was revealed that, although short reviews are pivotal for sentiment analysis, short reviews (word count ≤ 3) do not significantly contribute to the generation of well-definedand meaningful topics in topic modeling. The quality of the generated topics for UX factor identification was quantitatively evaluated using coherence scores. Scores of 0.56 and 0.49 were obtained for positive and negative topics, respectively, indicating the effectiveness of the topic generation process. In addition, word embedding was utilized to prioritize the topics generated from topic modeling. There, the thumbs-up count of the reviews playsa significant role in identifying the most influencing UX factorsof educational mobile apps. The proposed method serves as a guide for researchers and practitioners to extract and prioritize UX factors from mobile app reviews in various domains.**

***Keywords*—Mining app reviews, Topic modeling, Thumbs-up count, User experience, Education apps**

I. Introduction

Owing to the rapid advancement of mobile computing, daily human practices have become more efficient and convenient. Especially, due to the COVID-19 pandemic, human-to-human interactions, collaborative works and financial transactions were mostly happening through online platforms. This compelled individuals to rely on digital devices like smartphones for communication. Consequently, the mobile app market has witnessed a significant surge in the recent past, driven by the increasing demand for diverse user requirements. For instance, current statistics reveal that an average of 1,753 new apps have been released on the Google Play store per day [1]. This surge in app usage emphasizes the critical importance of developing highly reliable and user-friendly apps in this competitive app market. As such, is crucial to understand users' genuine experiences. Hence, popular platforms like the Google Play Store provide users with the opportunity to share their authentic experiences through user reviews, accompanied by an overall satisfaction rating, enabling the collection of valuable user feedback about the app UX. However, there are a number of instances where disagreements exist between user ratings and user reviews [2].

As such, analyzing user reviews gained significant attention [3]. For example, researchers have used various techniques such as Machine Learning (ML) to gain semantic meaning of the user reviews [4], [5]. Both supervised and unsupervised ML methods have been used in review analysis, but still, they have their own limitations. For instance, supervised methods rely on predefined labels, demanding substantial manual effort to create a ground truth dataset. This manual annotation process introduces limitations, including potential conflicts arising from the subjective nature of user opinions when interpreted by multiple annotators [6].

Nevertheless, Natural Language Processing (NLP) techniques, particularly topic modeling, offer an effective solution to uncover the themes embedded in reviews, addressing the limitations mentioned earlier. Among the various topic modeling algorithms, Latent Dirichlet Allocation (LDA) standsout as a widely employed method for extracting cohesive topics from reviews. Following the generation and interpretation of these topics from user reviews, the next crucial step involves identifying the most significant topics and prioritizing them. This prioritization is pivotal for making informed decisions regarding app enhancements. While previous research have focused on identifying topics in app user reviews across different domains, our knowledge indicates a gap in research that specifically addresses the prioritization of these topics. Therefore, in this research, we prioritize topics by calculating the total thumbs-up count for reviews associated with each topic in the generated list. This prioritization method allows us to interpret the most influential UX factors. For identifying relevant reviews, we employed the Word2Vec model for word embedding [7]. To validate our approach, we utilizeda user review dataset collected from three popular education apps, specifically recognized as Massive Open Online Course (MOOC) apps. With that in mind, we set the objectives of this research as follows:

- Assessing the difference between app ratings and sentiment scores derived from user reviews.
- Introducing a novel methodology to determine the most influential UX factors in mobile apps. In this case, we used the thumbs-up count to prioritize topics extracted from user reviews and considered a set of well-known educational mobile apps to test the proposed method.

II. RELATED WORK

App review mining is a widely used approach to gather authentic user feedback across various domains, aiming to optimize user experience (UX). Researchers employ supervised, unsupervised, and semi-supervised machine learning (ML) methods for different objectives such as topic identification, summarization, and review labelling [4], [6], [8]. Researchers have tackled several challenges in app review mining, addressing issues such as domain-specific challenges in topic modeling through seeding keywords [3] and automating the summarization of large-scale user reviews for detailed app comparisons [9]. Additionally, some studies have employed statistical techniques to quantitatively compare user sentiments with ratings [2], [10]. Furthermore, prior research has explored education app reviews using different approaches and objectives [11], [12]. For instance, one study delved into the analysis and comparison of user satisfaction with interactive education apps incorporating Augmented Reality and Virtual Reality compared to traditional educational apps [13].

A. *UX Factors in mobile apps*

UX factors are aspects related to the UX while using any product or design. In the context of the digital world, identifying the UX factors is crucial not only for app developers to develop usable apps that meet user requirements but also to focus on competitive advantage by giving pleasurable UX. It is vital to understand the factors which prominent to provide positive and negative UX so that the developers can improve the UX accordingly. For example, factors such as Bugs and Crashes, Improvement requests, Resource use, and Compatibility are associated with negative reviews [14].

III. METHODOLOGY

This section describes the experimental design and analysis employed in this study. A summary of the experimental process is shown in Figure 1. First, we removed duplicate reviews and cleaned the remaining reviews using text pre-processing techniques commonly used in NLP. Those are lowercasing, tokenization, removing punctuation and stop-words, and lemmatization, respectively. Both stemming and lemmatization are used to convert the words in the processed reviews into their root forms. Stemming achieves this by removing characters or letters from the original word, and hence, it may cause a loss of the semantic meaning of the word. In contrast, lemmatization converts the word into its dictionary form preserving the meaning [15]. Therefore, we adapted to lemmatization in this study, as the exact meaning of the word is essential for topic modeling.

Then, non-English reviews were removed and sentiment analysis was conducted using the Valence Aware Dictionary and sEntiment Reasoner (VADER), a lexicon and rule-based tool specifically designed for social media text [16]. VADER was employed for sentiment analysis, categorizing reviews into positive, negative, and neutral sentiments. The analysis proceeded in two main directions. First, a statistical analysis quantified the alignment between each user's review and their app rating. Second, influential UX factors were prioritized. Those two are explained in detail in the subsequent sections.

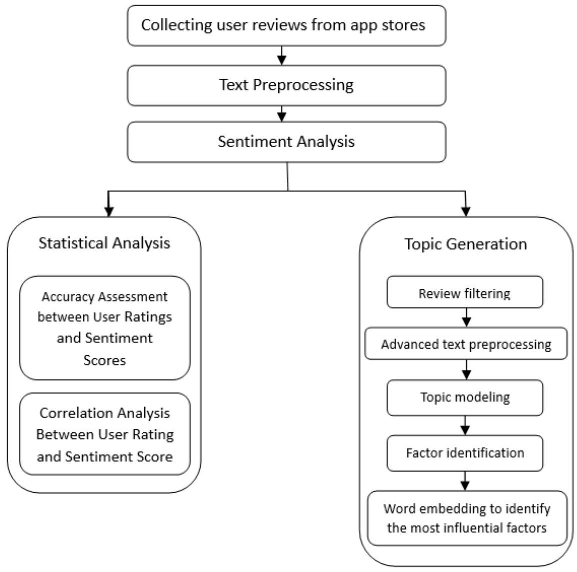

Fig. 1. Methodology of the paper

A. *Data and App Sample Selection*

As indicated by Statista [17], Education is the second most popular app category in the Google Play Store in 2022. Consequently, this study focuses on education app reviews, leveraging the popularity of this category. Therefore, three popular education apps were systematically selected based on the number of downloads and user reviews to evaluate the proposed method. Those apps are Udemy, Coursera, and edX. User review data of these apps were scraped from the Google Play Store using the open-source tool google_play_scarper. Each dataset contains 26088, 19354, and 17978 reviews respectively after eliminating duplicates.

B. *Statistical Analysis*

Our statistical analysis is two-fold. First, an accuracy assessment aimed at quantifying the disparity between the sentiment score and the corresponding user rating using Equation 1. For instance, an accuracy of 40% for a rating score of 2 implies that, out of a total of 150 reviews with a rating score of 2, 60 were correctly identified with a sentiment score of 2 during sentiment analysis.

$$Accuracy\ for\ Rating\ Score\ x = \left(\frac{Accurate\ reflection\ for\ rating\ score\ x}{Total\ data\ points\ for\ rating\ score\ x}\right) \times 100 \qquad (1)$$

Here, Accuracy represents the percentage of cases where the numeric rating accurately reflects the sentiment in the review and

- Accurate reflections for rating score x: Instances where the user rating matches the sentiment score derived from the review.
- Total data points for rating score x: The total number of instances or data points for each rating category.

Second, correlation analysis was conducted to know whether a relationship exists between users' reviews which reflect their subjective opinions, and user ratings. This analysis provides insights into how users' opinions align with user satisfaction quantified as ratings. Pearson's correlation

TABLE I. CUSTOM STOPWORDS REMOVED FROM THE REVIEWS

Type	Custom stop words
Domain-specific (Education) terms	'education', 'learn', ' learning', ' student', 'teach', 'teaching'
App-specific terms	'edx', 'udemy', 'coursera', 'app', 'application'
Common adjectives	'amazing', 'good', 'bad', 'really', 'awesome', 'great', 'enjoy', 'wonderful', 'love', 'best', 'excellent', 'nice', 'easy'
Common verbs	'use', 'try', 'like', 'could', 'get', 'through'
Greetings	'thank', 'thanks', 'may'

coefficient was used as it is suited for assessing linear correlation with two quantitative variables. The correlation coefficient value ranges between -1 and 1. If the value is below ± 0.4, a low correlationis reflected; above ± 0.6 shows a high correlation, and between ± 0.4 and ± 0.6 shows a moderate correlation [10].

C. Topic Generation

We focused on identifying topics in negative and positive reviews, as neutral reviews lack usefulness unless subjected toin-depth analysis for sentiment categorization. During the topic generation phase, reviews containing three words or fewer were initially filtered out due to their limited contribution to identifying UX factors. Custom stop words and emojis were subsequently removed during advanced text preprocessing, as they typically do not contribute to topic modeling. The list of removed custom stop words is presented in Table I.

The LDA algorithm was employed for topic modeling on the processed dataset, with the specific goal of distinguishing topics within positive and negative reviews. The implementation utilized an LDA model on a preprocessed sentence corpus created by combining all app reviews from the three mentionedapps. The resulting dataset for identifying influencing factors encompassed a total of 39,550 user reviews. To evaluate the model's performance, Perplexity and Coherence Score metrics were employed. Perplexity measures the model's predictive likelihood, while the Coherence Score assesses the meaningfulness of topics [18]. Fine-tuning of hyper-parameters, including alpha and beta values, as well as the number of topics, was conducted to optimize model performance. Finally, the generated topics were categorized and aligned with UX factors [14] to pinpoint the factors influencing UX in education mobile apps.

Moreover, some of the worthy previous research has considered the thumbs-up count as a helpfulness rating for the reviews and considered only those with at least one thumbs-up count for their analysis, aiming to enhance model accuracy. However, some reviews crucial for app improvement have not received a thumbs-up so, the count of thumbs-up shows as 0. This is due to the fact that the users might not read all reviews. Thus, a limitation arises as filtered-out reviews might contain valuable information essential for identifying UX factors specific to the app [19]. In our dataset, filtering reviews with at least one thumbs-up count resulted in excluding more than 50% of reviews for each chosen app. Table II illustrates the number of reviews with thumbs-up counts separately categorized as positive and negative, along with the total number of thumbs-up counts

TABLE II. THE NUMBER OF REVIEWS WITH THUMBS-UPS

App	As a percentage	Number of positive reviews	Number of negative reviews
Udemy	7.67%	1199	802
Coursera	26.10%	3587	1465
edX	26.04%	2426	330

represented as a percentage.

Yet, thumbs-up counts signify the level of agreement among the app users and thus increase the reliability of the information given by a user. Therefore, the thumbs-up count in app reviews is useful for prioritizing the UX factors. For example, in Table III, three negative reviews are presented along with their respective thumbs-up counts. The first review raises an issue with 27 thumbs-ups, indicating a widespread concern among users, while the third review with 0 thumbs-ups suggests a less prevalent problem. This discrepancy alerts app designers to pay more attention to the sign-in issue mentioned in the first review due to its higher user consensus, thus highlighting the reliability of its content.

Therefore, prioritizing UX factors is facilitated by considering the thumbs-up count in app reviews in this research. For this purpose, after generating the topics, a Word2Vec model was trained on the tokenized reviews to identify those that include the generated topics. For each word in the generated topic list, three similar words were identified using the Word2Vec embeddings. Positive and negative thumbs-up counts were then calculated for each topic, considering the presence of at least two similar words within each topic in a given review. This process allowed for the prioritization of identified topics, which were subsequently mapped to UX factors extracted from existing literature [14]. The results of this prioritization will be discussed in the next section.

IV. RESULTS AND DISCUSSION

This section presents and discusses the results of the statistical analysis and topic generation. The overall accuracy percentage in the accuracy assessment, as shown in Table IV, is found to be less than 40%. Separate assessments were conducted for each rating category, both with and without a threshold of 1, where a threshold of 1 indicates sentiment scores considered accurate within 1 unit of the actual rating. However, some categories, like Udemy app ratings 1 and 2, showed accuracy percentages (34.29% and 64.92%) below expectations, suggesting user reviews may differ from actual ratings even with the threshold.

In correlation findings, Pearson coefficients for Udemy, Coursera, and edX datasets are 0.52, 0.45, and 0.33 with p-values of 0.000, indicating statistically significant relationships. The edX dataset exhibits a weak positive correlation, while the other two show a moderately positive correlation between app ratings and review scores. Addressing the first research objective, assessing the accuracy of sentiment scores against user ratings as ground truth reveals variations, emphasizes the need for advanced text analysis to identify detailed user suggestions, requests, and issues for UX enhancement.

TABLE III. SAMPLE OF NEGATIVE REVIEWS WITH THUMBS-UP COUNTS

Review	**thumbs-ups**
Can't sign in with Facebook. It doesn't go beyond sign in screen. It always stays on sign in screen even though I'm already signed in	27
I am having problems with quizzes on the app when they contain pictures. It doesn't load the multiple choice answers and the submit button after the picture.	7
Cannot open pdf file links	0

TABLE IV. ACCURACY OF SENTIMENT SCORES AGAINST USER RATINGS

	edX	**edX with threshold 1**	**Udemy**	**Udemy with threshold 1**	**Coursera**	**Coursera with threshold 1**
All ratings	38.12%	74.42%	32.47%	68.72%	33.54%	70.81%
Rating score 5	40.52%	73.52%	49.05%	75.37%	40.65%	74.33%
Rating score 4	39.76%	96.59%	43.66%	96.00%	33.77%	94.61%
Rating score 3	37.07%	79.21%	36.72%	80.18%	36.88%	79.89%
Rating score 2	16.33%	66.93%	19.57%	64.92%	18.68%	66.02%
Rating score 1	06.88%	31.04%	11.55%	34.29%	09.15%	32.44%

A. *Topic Identification*

This section discusses the outcomes of topic generation and UX factor identification. Table V shows review counts after each preprocessing stage. The dataset size remained largely consistent after fundamental steps like punctuation and stop word removal. However, a notable reduction occurred during short review filtering, primarily due to the prevalence of one or two-word reviews lacking meaningful information for topic identification.

During LDA model hyper-parameter tuning, optimal alpha values were found to be 0.5 for positive reviews and 0.3 for negative reviews, with a consistent beta value of 0.3. Figure 2 depicts the distribution of the number of topics against the coherence score. The visualization indicates an ideal number of 5 topics for positive reviews and 7 topics for negative reviews. Notably, coherence scores consistently increased from topic 2 to 7 for negative reviews, suggesting improved interpretability. However, a sharp decline from 0.48 to 0.38 beyond 7 topics indicates diminishing returns. Table VI shows coherence scores and perplexity after each filtering stage, revealing a progressive improvement and well-defined topics with closely associated words.

TABLE V. THE NUMBER OF REVIEWS AFTER EACH FILTERING STAGE

App	**Initial**	**After removing short reviews (≤ 3)**	**Custom stop words removal**
Udemy	26088	17932	17071
Coursera	19354	13056	12759
edX	17978	10668	9720

TABLE VI. COHERENCE SCORE AND POLARITY AFTER EACH FILTERING STAGE

	Positive reviews		Negative reviews	
	Coherence score	Perplexity	Coherence score	Perplexity
Removing custom stop word	0.53	-7.502	0.38	-7.055
Removing shortreviews	0.55	-7.46	0.44	-7.09
hyper-parameter tuning	0.56	-7.46	0.49	-7.13

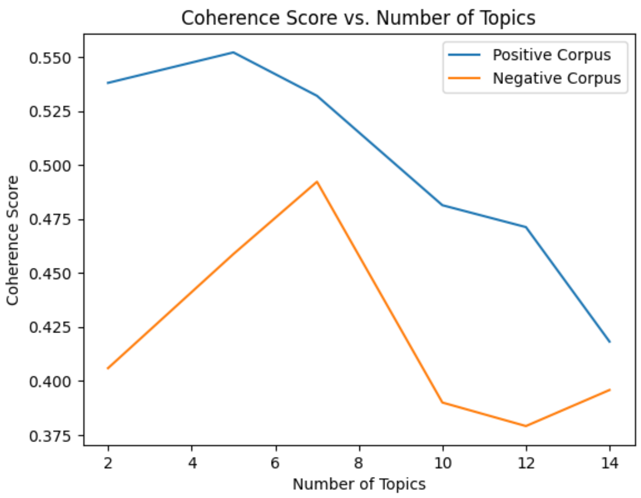

Fig. 2. Coherence score vs. number of topics

The final results of topic identification and UX factor mapping are presented in Table VII, including the generated topics, example reviews, topic interpretations, corresponding UX factors, and thumbs-up counts for each topic. In this context, a positive thumbs-up count signifies the count from positive reviews, while a negative thumbs-up count represents the count from negative reviews based on the sentiment of each review. These topics were associated with UX factors identified in previous literature [14], [20]. Notably, Topics 9, 10, 2, 6, and 7 received the highest thumbs-up counts in descending order, highlighting concerns related to login issues, accuracy and accessibility in the quiz feature, cross-device compatibility, video playback speed issues, and the significance of helpful user guidance and a user-friendly interface in the context of the selected MOOC apps based on the dataset. Notably, despite a lower count, Topic 1 underscores the significance of information hierarchy in MOOC app design. This analysis successfully achieved the second research objective by prioritizing impactful UX factors from app reviews. These findings guide education app developers to prioritize impactful aspects and address crucial issues in designing and improving apps.

In summary, this research proposed a methodology to identify influential UX factors from mobile app reviews. The study emphasized the significance of thumbs-up count in prioritizing these factors through the generated topics which is not considered by the previous similar studies [3],[11],[13].

TABLE VII. MAPPING TOPICS WITH EXISTING UX FACTORS

	Topic	Prominent words	Example review	Topic interpretation	UX factor/s	Thumbs-ups counts (Positive, Negative)
Positive	Topic 1	'content', 'well', 'way', 'quality', 'used', 'information', 'make', 'place'	Coursera is best of all MOOC platforms. It provides high quality study materials from top universitiesalong with flexibility to learn at own pace.	Well-structured content	Information architecture/Ease of use	22, 9
	Topic 2	'help', 'much', 'option', 'go', 'user', 'recommend', 'interface', 'friendly'	Reliable access to resources through the app in my experience. Very useful to watch course content on the go and the ability to download and watch offline.I often also watch lectures on my phone so I can work along with it on my computer. Good stuff.	User friendly interface/ User-centric design	Ease of use/ Attractiveness/ Interface	6763, 3739
	Topic 3	'course', 'platform', 'free', 'knowledge', 'skill', 'study', 'new', 'university'	This has to be the coolest app. Great classes, andthe opportunity to learn whatever you want for free. Give thanks	Diverse free courses	Content variety/ Cost/ Comparison	5668, 689
	Topic 4	'need', 'better', 'available', 'always', 'everyone', 'improve', 'language'	Why does the subtitles can't show up and keeps saying that error. So I hope you make better improvement and thank you for this app	Enhancement Requests	Improvement request/ Feature/ Functionality	1851, 271
	Topic 5	'video', 'content', 'play', 'screen', 'phone', 'version', 'button', 'useful'	I would prefer touching screen anywhere only once to pause or resume, because I had to pause these videos frequently	Usability of buttons and controls in the video feature	Feature (Video)	2540, 2513
Negative	Topic 6	'quit', 'video', 'long', 'option', 'start', 'begin', 'time', 'model'	Add possibility to play video lectures 1.25, 1.5, 1.75, 2.0 times faster	Video Play issues	Performance in video feature	6637, 3457
	Topic 7	'issue', 'android', 'device', 'display', 'hate', 'kind', 'often', 'coming'	Wanted to do my coursework on my Kindle Fire tablet, but this app doesn't support DIRECTLY fromAmazon. Coursera staff were kind enough to direct me to a YouTube video that shows how to get aroundthis. It seems kind of silly to have to dance around like this to do something routine, but I am happy now!	Device compatibility issues	Cross-Device compatibility	5630, 3877
	Topic 8	'crash', 'open', 'give', 'install', 'error', 'expert', 'bug', 'fix'	After playing the video..within a minute applicationget crashes..please look into this asap	Technical issues and errors	Bugs/ Crashes/ Customer support	801, 719
	Topic 9	'issue', 'log', 'start', 'login', 'still', 'account', 'password', 'sign'	It seems great, and a great reviews, but I can't even get passed the login. It keeps returning as though I refreshed it.	Login issues	Bugs/ Feature (Login)	6365, 7160
	Topic 10	'way', 'answer', 'submit', 'question', 'quite', 'quality', 'material', 'test'	I'm not being able to take the tests. Every time I try for tests, a message stating, "This quiz is currently only supported in web" pops up.	Problems in accessing materials	Accuracy and Accessibility (Quiz feature)	6648, 7079
	Topic 11	'frustration', 'find', 'terrible', 'done', 'disappointing', 'chromecast', 'interface', 'would'	Much better now that they have Chromecast support,still a little buggy but much better.	Negative Chromecast UX	Feature improvement/ Bugs	2902, 2640
	Topic 12	'much', 'user', 'come', 'need', 'dark', 'mode', 'mobile', 'study'	Why is there no dark mode? Holy hell your app isburning my retinas.	Request for dark mode	Feature request (Dark mode)/ Personalization	1387, 988

To address the limitation of recent reviews lacking thumbs-up counts, the study evaluated the method using the latest reviews and recommends applying this method for review datasets collected in a specific time period. A limitation arises from interpreting usability concerns collectively for all three apps. However, while some studies do not extend to identify UX factors from generated topics [3] and others interpret topics without prioritizing UX factors [12], this study successfully introduces thumbs-up counts for prioritization, identifying UX factors in the education domain across three MOOC apps. Background information like posting year and app version wasn't explicitly considered. Future research can extend this work by addressing these limitations and automating the processes of topic interpretation and UX factor mapping, which were carried out manually in this study.

V. CONCLUSION

This study introduces a novel methodology for analyzing and prioritizing UX factors in mobile app user reviews, utilizing thumbs-up counts through the application of topic modeling and word embedding. Firstly, the disparity between user ratings and corresponding sentiment scores was quantified. Secondly, the UX factors were ranked to identify the most influential factors from user reviews. The topics were then interpreted and mapped with UX factors from the literature. The method was evaluated using a dataset of 63,420 user reviews from prominent MOOC mobile apps (Udemy, Coursera, and edX) in the education app domain. A key finding highlights the importance of the thumbs-up count as a critical measure in prioritizing UX factors. These results are useful foroptimizing the UX in mobile apps. Moreover, this method canbe effectively utilized to identify the most influential factorsof mobile apps in other domains as well.

REFERENCES

[1] Google Play Statistics and Trends 2023, December 05, 2023 (Accessed: December 5, 2023). [Online]. Available: https://42matters.com/google-play-statistics-and-trends#apps-released-per-day

[2] P. Lak and O. Turetken, "Star ratings versus sentiment analysis – a comparison of explicit and implicit measures of opinions," *Proceedings of the 2014 47th Hawaii International Conference on System Sciences,* pp. 796–805, Jan. 2014, doi: 10.1109/HICSS.2014.106.

[3] M. Tushev, F. Ebrahimi, and A. Mahmoud, "Domain-specific analysisof mobile app reviews using keyword-assisted topic models," in *2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE)*, pp. 762–773, 2022, doi: 10.1145/3510003.3510201.

[4] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Analysing app reviews for software engineering: a systematic literature review," *Empirical Software Engineering*, vol. 27, pp. 1–63, 2022, doi: 10.1007/s10664-021-10065-7.

[5] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora, and H. C. Gall, "What would users change in my app? summarizing app reviews for recommending software changes," *in Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, ser. FSE 2016, New York, NY, USA: Association for Computing Machinery*, pp. 499-510, Nov. 2016, doi: 10.1145/2950290.2950299.

[6] V. T. Dhinakaran, R. Pulle, N. Ajmeri, and P. K. Murukannaiah, "App review analysis via active learning: Reducing supervision effort without compromising classification accuracy," in *2018 IEEE 26th InternationalRequirements Engineering Conference (RE)*, pp. 170–181, Oct. 2018, doi: 10.1109/RE.2018.00026.

[7] M. Fauzi, "Word2vec model for sentiment analysis of product reviews in indonesian language," *International Journal of Electrical and ComputerEngineering (IJECE)*, vol. 9, p. 525-530, July 2018, doi: 10.11591/ijece.v9i1.pp525-530.

[8] T. Shaik, X. Tao, Y. Li, C. Dann, J. McDonald, P. Redmond, and L. Galligan, "A review of the trends and challenges in adopting natural language processing methods for education feedback analysis," *IEEE Access*, vol. 10, pp. 56 720–56 739, 2022, doi: 10.1109/ACCESS.2022.3177752.

[9] Y. Li, B. Jia, Y. Guo, and X. Chen, "Mining user reviews for mobile app comparisons," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, Sep 2017, doi: 10.1145/3130935

[10] A. Finkelstein, M. Harman, Y. Jia, W. Martin, F. Sarro, and Y. Zhang, "Investigating the relationship between price, rating, and popularity in the blackberry world app store," *Information and Software Technology*, vol. 87, pp. 119-139, Mar. 2017, doi: 10.1016/j.infsof.2017.03.002.

[11] M. Haering, M. Bano, D. Zowghi, M. Kearney, and W. Maalej, "Automating the evaluation of education apps with app store data," *IEEE Transactions on Learning Technologies*, vol. 14, no. 1, pp. 16–27, Feb. 2021, doi: 10.1109/TLT.2021.3055121.

[12] S. Graser, L. H. Nielsen, and S. Bo¨hm, "Factors Influencing the User Experience of Mobile Augmented Reality Apps: An Analysis of User Feedback Based on App Store User Reviews." *Cham: Springer International Publishing*, pp. 109–129, Oct. 2023, doi: 10.1007/978-3-031-39940-4_9.

[13] A. S. Mondal, Y. Zhu, K. Bhagat, and N. Giacaman, "Analysing user reviews of interactive educational apps: a sentiment analysis approach," *Interactive Learning Environments*, June 2022, doi: 10.1080/10494820.2022.2086578.

[14] W. T. Nakamura, E. C. de Oliveira, E. H. de Oliveira, D. Redmiles, and T. Conte, "What factors affect the ux in mobile apps? a systematic mapping study on the analysis of app store reviews," *J. Syst. Softw.*, vol. 193, no. C, Nov. 2022, doi: 10.1016/j.jss.2022.111462

[15] D. Khyani and S. B S, "An interpretation of lemmatization and stemming in natural language processing," *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, vol. 22, pp. 350–357, Jan. 2021.

[16] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," *Proceedingsof the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216–225, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[17] L. Ceci, *Most popular Google Play app categories 2022*, Aug 23, 2023 (Accessed: December 2, 2023). [Online]. Available: https://www.statista.com/statistics/279286/google- play-android-app-categories/

[18] A. Fan, F. Doshi-Velez, and L. Miratrix, "Assessing topic model relevance: Evaluation and informative priors," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 12, no. 3, pp. 210–222, 2019, doi: 10.1002/sam.11415.

[19] A. W. de Jong, "Making Sense of App Reviews: Efficient Analysis of User Reviews for Mobile Apps with STM," M.S. Thesis, Department of Information Systems, Production and Logistics Management, University of Innsbruck, Austria, Mar. 2022. [Online]. Available: https://diglib.uibk.ac.at/download/pdf/7617150.pdf

[20] A. Carvalhido, R. Novo, P. Faria, and A. Curralo, "A User Experience Design Process in Mobile Applications Prototypes: A Case Study", *Advances in Design and Digital Communication II*, pp. 262–273, Jan. 2022, doi: 10.1007/978-3-030-89735-2_22.