International Journal of Advanced Smart Convergence Vol.9 No.4 124-131 (2020)
http://dx.doi.org/10.7236/IJASC.2020.9.4.124

IJASC 20-4-16

User Review Mining: An Approach for Software Requirements Evolution

Jee Young Lee

Adjunct Professor, Department of Software, SeoKyeong University, Korea
J.Ann.Lee@skuniv.ac.kr

Abstract

As users of internet-based software applications increase, functional and non-functional problems for software applications are quickly exposed to user reviews. These user reviews are an important source of information for software improvement. User review mining has become an important topic of intelligent software engineering. This study proposes a user review mining method for software improvement. User review data collected by crawling on the app review page is analyzed to check user satisfaction. It analyzes the sentiment of positive and negative that users feel with a machine learning method. And it analyzes user requirement issues through topic analysis based on structural topic modeling. The user review mining process proposed in this study conducted a case study with the a non-face-to-face video conferencing app. Software improvement through user review mining contributes to the user lock-in effect and extending the life cycle of the software. The results of this study will contribute to providing insight on improvement not only for developers, but also for service operators and marketing.

Keywords: *Intelligent Software Engineering, User Review Mining, User Satisfaction Analysis, Sentiment Analysis, Structural Topic Modeling.*

1. Introduction

Internet-based software applications have diversified into software that supports daily life as well as specialized work software due to the rapid increase of mobile Internet users. In particular, the world is now demanding "social distancing" and non-face-to-face living due to COVID-19. Therefore, as the number of users of internet-based software applications supporting non-face-to-face life explodes, problems in terms of functions and non-functions for software applications are rapidly being exposed. In this situation, the functional and non-functional improvement of the software application must also be made quickly. In particular, it has become important to quickly and accurately analyze the satisfaction, dissatisfaction, and issues felt by current users in order to improve the internet-based software application used by many unspecified people. App user review mining, which deals with the automated classification and analysis of app user reviews, has recently become an attractive topic in intelligent software engineering [1-5]. Marketplaces such as the Google Play store not only facilitate mobile application distribution, but also allow users to easily rate and post reviews for downloaded apps. These reviews serve as a source of information for

Manuscript Received: October. 28, 2020 / Revised: November. 9, 2020 / Accepted: November. 13, 2020
Corresponding Author: J.Ann.Lee@skuniv.ac.kr
Tel: +82-2-940-7520
Adjunct Professor, Department of Software, SeoKyeong University, Korea

app developers to understand user needs and improve satisfaction [6].

Among review mining techniques, research using machine learning and natural language processing technology is attracting attention [1,7]. In the previous study, there was a study of clustering app reviews by using topic modeling based on Latent Dirichlet allocation (LDA), an unsupervised learning method [8]. There was a study on how to classify reviews into positive and negative sentiment, and provide review clusters derived by applying LDA to software developers [9]. There was a study that it was more effective to analyze the rating score together for user review analysis [10]. However, these previous studies have not performed a multifaceted analysis of user review data.

This study proposed a systematic user review mining method for software improvement. In addition, a case study was conducted with “ZOOM Cloud Meetings”, a video conferencing application that has rapidly increased in use due to non-face-to-face life in the COVID-19 situation. This study analyzes the sentiment and topic of reviews. And by analyzing the rating score together, it analyzes how the user's emotional experience is. Analyzing both user experience and functional/non-functional improvement requirements, it aims to provide insight on improvement not only for developers, but also for service operators and marketing.

2. Methods

The overview of the user review mining method proposed in this study is shown in Figure 1. In summary, after crawling and collecting user reviews in Google Play, user satisfaction and dissatisfaction were quantified based on the rating score. By applying a machine learning method, the emotions experienced by the user were classified into positive and negative and analyzed. In order to derive and analyze user requirements issues, the structural topic model (STM), a topic model that improved the LDA model, was used.

For the implementation and analysis of the user review mining methodology proposed in this study, Python 3.8, pandas 1.1.4, machine learning analysis packages genism and scikit-lean, and Korean natural language processing package KoNLPy were used. In addition, STM package was used to derive user requirements issues [11].

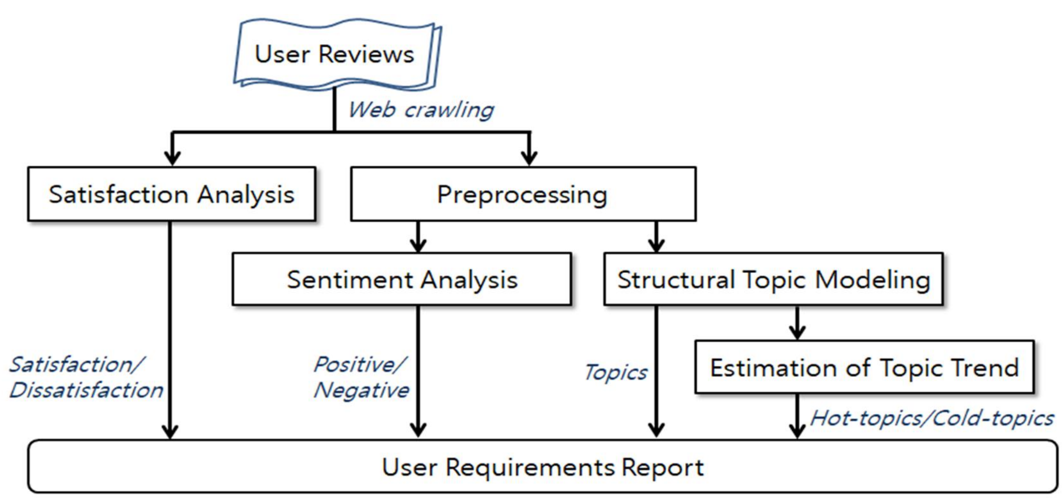


Figure 1. The overview of proposed user review mining method

2.1 User review data collection

User review data for the app was collected by crawling from the user review page in the Google play store. The Google play store provides users with a user review page where users who download, install and use the app can easily and quickly post reviews. Therefore, review contents, rating score, and creation date are collected using the user review page of the Google play store as a data collection source. The collected review content performs pre-processing. The mecab is used for korean morpheme analysis.

2.2 User satisfaction analysis

Satisfaction and dissatisfaction were classified based on the rating score in order to analyze user satisfaction with the app. If the rating score is 4 or higher, it is classified as satisfactory, and the rest are classified as unsatisfactory. Then, word frequency analysis is performed on the classified data. Based on the word frequency, it is vectorized into a bi-gram to compare and analyze the top words of satisfaction and dissatisfaction reviews.

2.3 User sentiment analysis based on machine learning

Sentiment classification analysis is performed to identify the positive and negative sentiment the user has with the app. Among the machine learning methods, logistic regression analysis is performed to generate a sentiment classification model. For the training of the sentiment classification model, the NAVER movie sentiment data, an open dataset, is used.

2.4 STM-based user review issue analysis

STM is a topic model that improves and extends LDA. STM determines the distribution of words constituting a topic using not only the frequency of words in the document but also the document metadata (author's gender, age, year, etc.) [12]. As shown in Figure 2, STM can be conceptually divided into three components; (1) a topic prevalence model, which controls how words are allocated to topics as a function of covariates, (2) a topical content model, which controls the frequency of the terms in each topic as a function of covariates, and (3) a core language model which combines these two sources of variation to produce the actual words in each document [13-14].

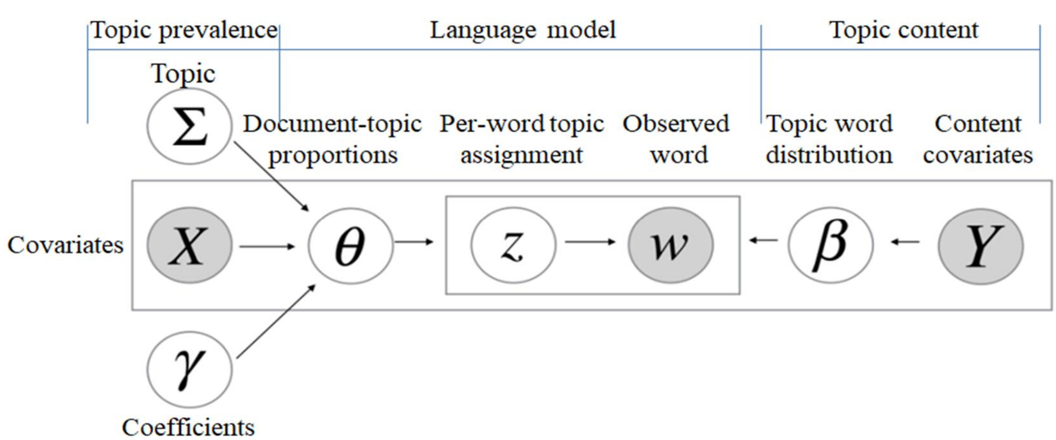


Figure 2. Conceptual structure and process of STM

For topics derived by SMT, hot-topic which is an uptrend and cold-topic which is a downtrend are identified by setting the rating scale and date as covariates. Next, since STM uses the covariance matrix of the logistic normal distribution, the correlation between topics is estimated.

3. Results and discussion

3.1 User review data collection

As of October 31, 2020, there were 1,184,859 user reviews on the Zoom Meeting review page of the Google play store. In the review contents, excluding numbers and symbols, those with more than 20 letters were crawled under filtering conditions, and 1,402 were collected as research data.

3.2 User satisfaction analysis

Satisfaction and dissatisfaction classification were performed based on the rating score of the user review. Satisfaction and dissatisfaction were classified based on the user review rating score. Figure 3 shows the ratio of satisfaction reviews and dissatisfaction reviews on a monthly basis. There were no user reviews in May 2020, and only satisfactory reviews in April. From July to September, the percentage of dissatisfied reviews increased.

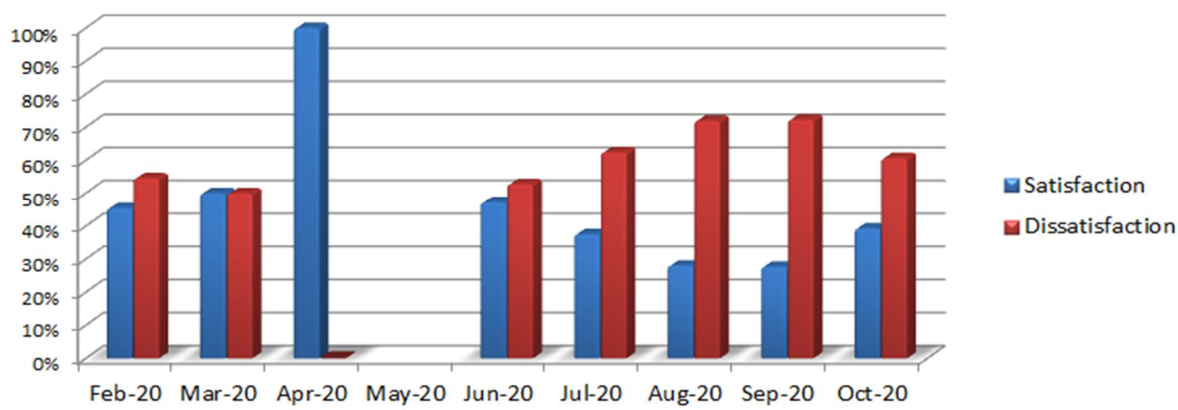


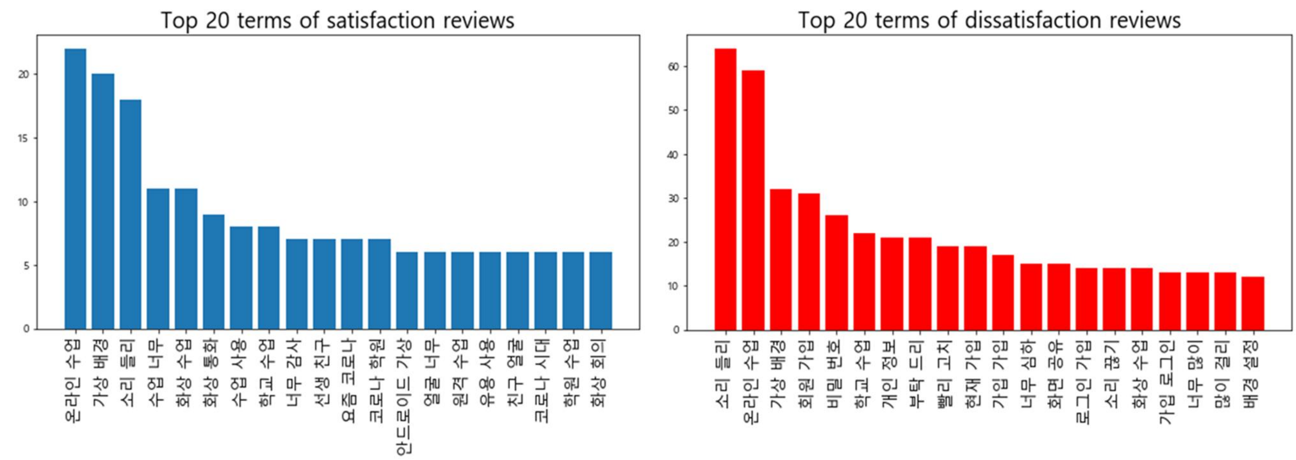
Figure 3. Ratio of satisfaction reviews and dissatisfaction reviews by month

Figure 4 is a graph that analyzes the word frequency for the content of satisfaction and dissatisfaction reviews in a bi-gram method. Looking at the top words of the satisfaction review in Figure 4-(a), there are 'online class', 'remote class', and 'friend face'. In the non-face-to-face life caused by the COVID-19, it was useful to be able to do online classes through the Zoom Meeting app, and they were satisfied that they were able to see a friend's face. Looking at the top words in the dissatisfied review in Figure 4-(b), 'sound can be heard', 'online class', and 'membership registration' were found. There were dissatisfied reviews that the sound was cut off in online classes and that membership registration was uncomfortable.

3.3 User sentiment analysis based on machine learning

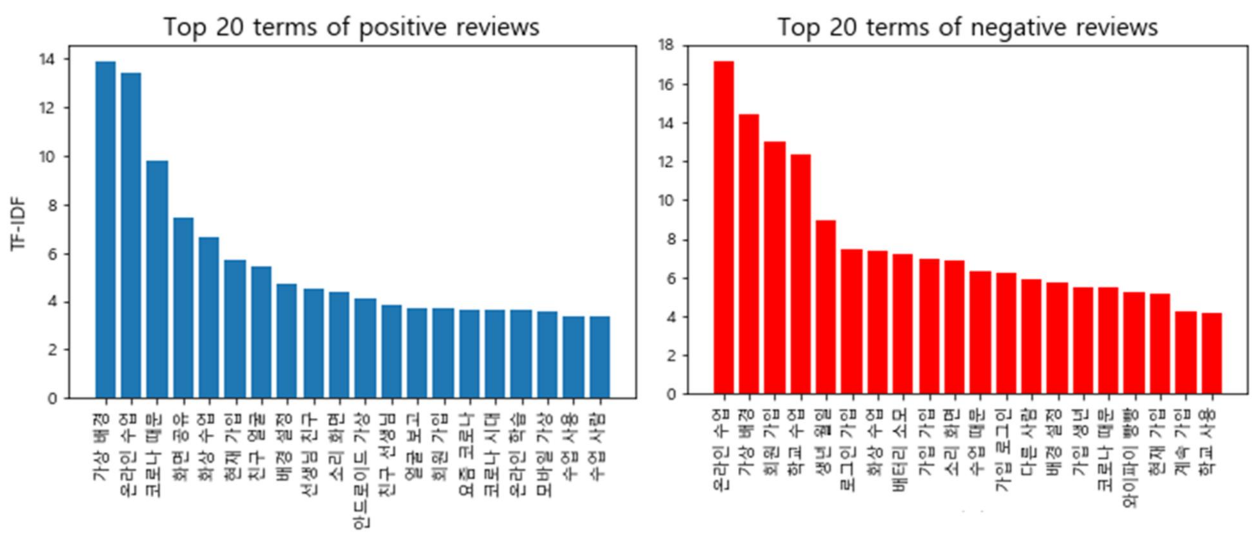
We classified sentiment by applying user review data to the sentiment classification model created and learned in section 2.3. Figure 5 shows the results of the sentiment classification analysis of positive and negative. Looking at the top words of the positive sentiment in Figure 5-(a), there are 'virtual background', 'online class', 'because of the corona', and 'face of friends'. There were positive sentiment about being able to see face of friends that they had not seen because of the corona while taking online classes.

In addition, looking at the top words for negative sentiment in Figure 5-(b), 'online class', 'virtual background', 'member sign-up', and 'battery consumption' were found. There were reviews saying that they did not want to take online classes and that the virtual background setting was poor, that it was difficult to sign up for membership, and that the battery was consumed.



(a) (b)

Figure 4. Word frequency on user satisfaction analysis: (a) Top 20 words in satisfaction reviews (b) Top 20 words in dissatisfied reviews



(a) (b)

Figure 5. Word frequency on sentiment classification analysis of user reviews: (a) Top 20 words in positive reviews (b) Top 20 words in negative reviews

3.4 STM-based user review issue analysis

In order to determine the optimal number of topics, 4 indicators such as held-out likelihood, residual, semantic coherence, and lower bound were identified while gradually increasing the number of topics from 5 to 15. Finally, we decided the optimal number of topics was 11.

By performing STM, it is possible to derive high-probability words for topic. Among those words, three words that were determined to be analyzed are summarized in Table 1. In addition, five user reviews with a high proportion of the topic were analyzed together with the high-probability words to determine label that describes the topic.

Table 1. STM based topic and top words extraction

Topic	Top words	Topic label
1	sound (소리), Inaudible (안들려요), break (끊겨서)	Sound problem
2	not working (안돼여), mobile data (모바일데이터), tablet (테블릿)	Mobile device
3	sign up (가입), date of birth (생년월일), none (없다)	Sign up
4	login (로그인), id (계정), password (비번)	Login
5	please (해주세요), participant (사람), host only (호스트만)	Participant authority
6	error (오류), entrance (들어가야하는데), meeting room password (회의에암호)	Meeting room password
7	mute (음소거), button (버튼), supportive sound (지지직소리)	Mute
8	school (학교), friends (친구들), lecture (수업)	School class
9	COVID-19 (코로나), home (집), meeting (모임)	Non-face-to-face meeting
10	virtual background (가상배경), android (안드로이드), settings (설정)	Change settings
11	class (수업), online (온라인), lecturer (선생님)	Online class

In this study, the covariate was set as a rating score and date for topics derived from STM, and the trend in topic proportions according to the writing date was analyzed. As shown in Figure 6, the decreasing cold-topic was Topic 2 (Mobile device) and Topic 9 (Non-face-to-face meeting). On the contrary, the increasing hot-topic appeared as Topic 10 (Change settings) and Topic 7 (Mute).

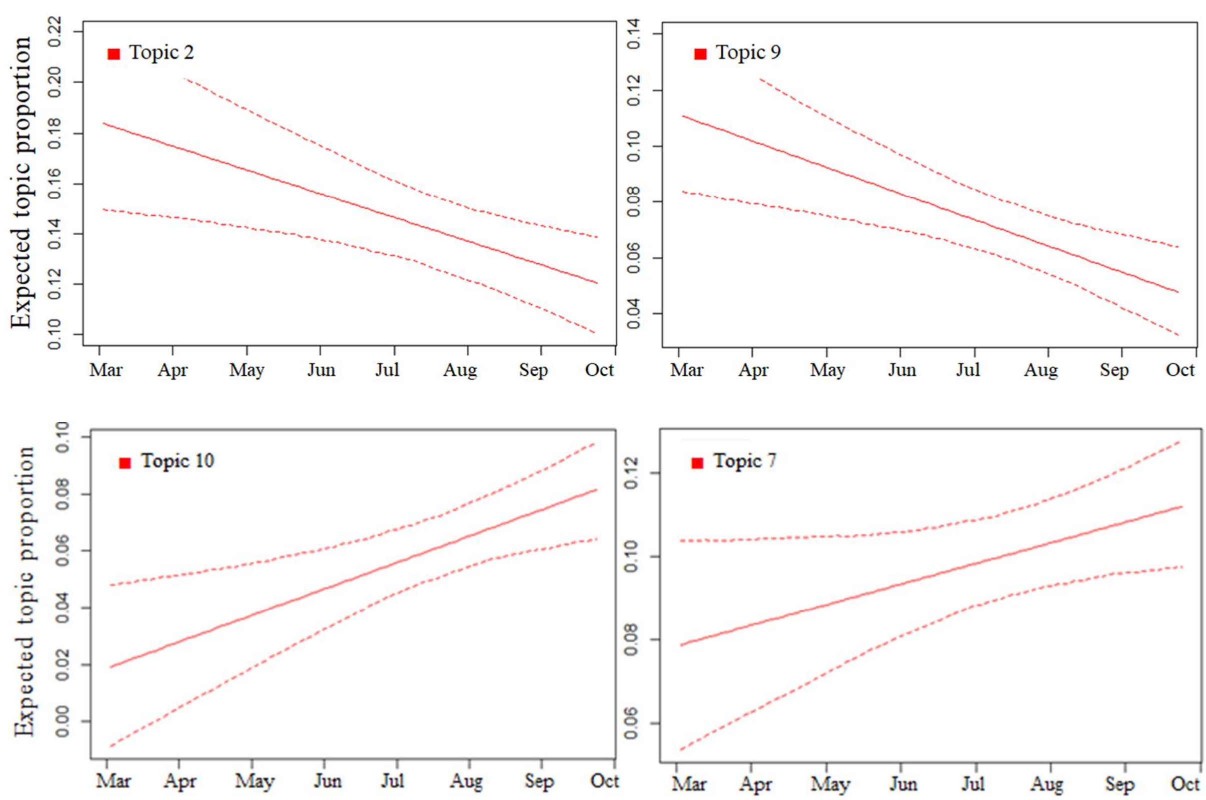


Figure 6. Estimation of topic trend by monthly

The trends in topics of user review can be analyzed to determine the importance of user requirements. The fact that Topic 2 is cold-topic seems to have improved the requirement that it is inconvenient to use in mobile devices compared to the PC version in the early days. Similarly, the fact that Topic 9 is cold-topic can be interpreted as a user accepting the adoption of non-face-to-face meetings. On the contrary, the hot-topic Topic 10 (Change settings) and Topic 7 (Mute) should be treated as user requirements that must be addressed urgently for software improvement.

In this study, there are 11 user requirements issues derived through topic modeling. It should be analyzed in conjunction with the results analyzed in 3.2 and 3.3. The user's satisfaction and sentiment experience could be grasped. The 'Zoom meeting' app used as an example in this study was introduced into non-face-to-face online classes and meetings from February 2020, and the number of users increased rapidly. It was analyzed that there were confusion and complain as the user suddenly used it involuntarily due to external factors without understanding how to use it.

4. Conclusion

This study proposed a user review mining method and process to improve internet-based software applications. The proposed method is to collect review data from the review page of the app store using a crawl method, perform a rating score-based user satisfaction analysis, user sentiment analysis using a machine learning-based sentiment model, and topic analysis of user requirement. The results of these analyses are feed-back to the development and/or the operations team as action items to improvement the software.

The limitation of this study is that the quality of user review data cannot be guaranteed. There were 1,184,859 reviews on the app review page of the Google Play Store, but only 1,402 had more than 20 characters that could be analyzed excluding numbers and symbols. And the lack of information on the user's age or software utilization level in the user review data is a limit to the accuracy of the analysis results.

The contribution of the user review mining method proposed in this study is as follows.

- Users' requirements can be immediately collected and analyzed.
- User loyalty to the software can be managed by analyzing user satisfaction.
- Through the analysis of the positive/negative emotion classification of the review text, it is possible to identify the requirements that users accept negatively, that is, needs to be improved.
- Topic analysis through STM can classify user requirements in detail, and among them, the requirements to be urgently improved can be identified.
- The proposed user review mining method can be applied to all software.

References

[1] M. Tavakoli, L. Zhao, A. Heydari, et al., “Extracting useful software development information from mobile application reviews: A survey of intelligent mining techniques and tools,” Expert Systems with Applications, vol. 113, pp. 186-199, 2018.
DOI: https://doi.org/10.1016/j.eswa.2018.05.037

[2] S. Panichella, A. Di Sorbo, E. Guzman, et al., "How can i improve my app? classifying user reviews for software maintenance and evolution," in 2015 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 281-290, IEEE, 2015.
DOI: https://doi.org/10.1109/ICSM.2015.7332474

[3] C. Tao, H. Guo, Z. Huang, “Identifying security issues for mobile applications based on user review summarization,” Information and Software Technology, vol. 122, pp. 106290, 2020.

DOI: https://doi.org/10.1016/j.infsof.2020.106290

[4] A. Ciurumelea, A. Schaufelbühl, S. Panichella, et al., "Analyzing reviews and code of mobile apps for better release planning," in 2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER), pp. 91-102, IEEE, 2017.
DOI: https://doi.org/10.1109/SANER.2017.7884612

[5] L. Villarroel, G. Bavota, B. Russo, et al., "Release planning of mobile apps based on user reviews," in 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE), pp. 14-24, IEEE, 2016.
DOI: https://doi.org/10.1145/2884781.2884818

[6] P. M. Vu, T. T. Nguyen, H. V. Pham, et al., "Mining user opinions in mobile app reviews: A keyword-based approach (t)," in 2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 749-759, IEEE, 2015.
DOI: https://doi.org/10.1109/ASE.2015.85

[7] J. Lee, “A Study on Research Trend Analysis and Topic Class Prediction of Digital Transformation using Text Mining,” International journal of advanced smart convergence, vol. 8, no. 2, pp. 183-190, 2019.
DOI: https://doi.org/10.7236/IJASC.2019.8.2.183

[8] N. Chen, J. Lin, S. C. Hoi, et al., "AR-miner: mining informative reviews for developers from mobile app marketplace," in Proceedings of the 36th international conference on software engineering, pp. 767-778, 2014.
DOI: https://doi.org/10.1145/2568225.2568263

[9] S. Moghaddam, "Beyond sentiment analysis: mining defects and improvements from customer feedback," in European conference on information retrieval, pp. 400-410, Springer, 2015.
DOI: https://doi.org/10.1007/978-3-319-16354-3_44

[10] W. Maalej, Z. Kurtanović, H. Nabil, et al., “On the automatic classification of app reviews,” Requirements Engineering, vol. 21, no. 3, pp. 311-331, 2016.
DOI: https://doi.org/10.1007/s00766-016-0251-9

[11] M. E. Roberts, B. M. Stewart, D. Tingley, “stm: R package for structural topic models,” Journal of Statistical Software, vol. 10, no. 2, pp. 1-40, 2014.
DOI: https://doi.org/10.18637/jss.v091.i02

[12] M. E. Roberts, B. M. Stewart, D. Tingley, et al., "The structural topic model and applied social science," in Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, Harrahs and Harveys, Lake Tahoe, 2013.

[13] M. E. Roberts, B. M. Stewart, E. M. Airoldi, “A model of text for experimentation in the social sciences,” Journal of the American Statistical Association, vol. 111, no. 515, pp. 988-1003, 2016.
DOI: https://doi.org/10.1080/01621459.2016.1141684

[14] J. Y. Lee, “A Study on Socio-technical System for Sustainability of the 4th Industrial Revolution: Machine Learning-based Analysis,” International Journal of Internet, Broadcasting and Communication(IJIBC), Vol.12, No.4, pp. 204-211, 2020.
DOI: https://doi.org/10.7236/IJIBC.2020.12.4.204