

Analysis of Success Factors of Electric Scooter Sharing Service Using User Review Text Mining

Kyoung-ae Seo* · Jung Seung Lee**

Abstract

This study aims to analyze service improvement and success factors of electric scooter sharing service companies by using text mining after collecting reviews of shared electric scooter service applications among various models of sharing economy. In this study, the factors of satisfaction and dissatisfaction of service users were identified using the term frequency inverse document frequency (TF-IDF) technique, and topics for each keyword were extracted using the Latent Dirichlet Allocation (LDA) Topic Modeling technique. According to the analysis results, the main topics were entertainment, safety, service area, application complaints, use complaints, convenience, and mobility. Using the analysis results of this study, employees and researchers of electric scooter sharing service companies will be able to contribute to the improvement and success of related services.

Keywords : Electric Scooter, Text Mining, Topic Modeling, TF-IDF, Word2Vec

Received : 2023. 02. 05. Revised : 2023. 02. 22. Final Acceptance : 2023. 02. 24.

* First Author, Doctoral candidate, Department of Business Administration, Hoseo University, e-mail: kacomssam@naver.com

** Corresponding Author, Associate Professor, School of Business, Hoseo University, 12, Hoseodae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do, 31066, Republic of Korea, Tel: +82-41-560-8361, e-mail: jslee@hoseo.edu

1. Introduction

As smart cities are spreading as a solution to problems caused by rapid urbanization, 'Smart Mobility' is naturally becoming more important [Kim et al., 2021]. Smart mobility is a new transportation service that combines smart devices and transportation services, and can be called a new transportation system [Hong and Park, 2011]. As the mobility environment changes, the requirements of shared mobility users are also increasing, and the connection service between transportation means to save users' travel time is also becoming increasingly important. Shared bicycles and shared electric scooters are representative means of shared mobility as means of moving from public transportation pick-up and drop-off points to destinations. As the number of users of shared electric scooters is rapidly increasing, research on this is also being actively conducted. Existing studies were sample research using questionnaires [Lee and Kim [2021], text mining research using news articles [Ryu and Cho, 2021], and topic modeling research using public data [Kim and Chung, 2006].

However, text mining analysis, which is a study that identifies users' intentions by directly extracting extensive review data using reviews written by users of shared electric scooter services, is still incomplete.

This study collects online review data written by users who used the electric scooter sharing service of the Play Store, analyzes the reviews, extracts key keywords, extracts positive and negative words, and classifies them by keyword. The purpose of this study is to examine the social issues and consumer awareness of the electric scooter sharing service by using the contents, and to suggest the

direction of improvement and development of electric scooter sharing service companies and electric scooter devices.

This study limited the scope to domestic sharing electric scooter service applications among several business models of the sharing economy. 'Wise App-App Analysis/Retail Analysis' In August 2021, four companies were selected among the electric scooter sharing service companies in the sharing service app rankings to collect online reviews. Among the top service providers, text mining analysis was conducted after collecting review data, excluding sharing service companies whose electric scooter sharing service was launched after 2020 or whose number of reviews was less than 2,500.

2. Related Research

2.1 Sharing Economy

The sharing economy refers to a consumption culture in which transportation means, lodging, places, and goods are shared with others, and not in possession of goods, but in cooperative consumption in which products or goods to be used necessary or selectively are shared. It can be said to be a model that creates new value by sharing objects or spaces owned by individuals or companies [Lessing, 2008].

People have come to enjoy a convenient and affluent life by sharing products or goods rather than sharing knowledge, and using social and human networks as a concept of sharing rather than ownership. cooperative consumption using information technology is being achieved. Recycling of resources reinforces interconnectedness between people and revitalizes the local economy, and shared serv-

ices have contributed greatly [Sundararajan, 2007].

As sharing economy services become more active, various studies are being conducted academically as well. Looking at related research cases, there is a study [Kang et al., 2015] that uses user reviews of the accommodation sharing platform Airbnb to extract words with high frequency and extract topics to identify users' intentions, and text mining to identify academic trends related to the sharing economy. There are studies that have identified the flow of related research using frequency analysis and topic modeling [Kim et al., 2021].

2.2 E-Scooter Sharing

The electric scooter sharing service refers to a service that uses an electric scooter, which is one of personal mobility devices that can run on electricity using an electric motor, for the purpose of renting for a short period of time. Unlike shared bicycles, there is no designated place for rental and return, and it is characterized by being able to return at any location within the service area [Lee et al., 2019].

As the shared electric scooter service began to be activated, various studies were conducted academically. There is a study that the use of shared electric scooters reduces the number of additional COVID-19 confirmed cases, or when the temperature is over 20 degrees, the use of electric scooters increases and decreases otherwise [Ryu and Cho, 2021]. In addition, there are studies proposing solutions for establishing return zones, introducing GPS systems, preventing privatization of bicycles, and using helmets to improve the citizenship of electric scooter users [Park, et al., 2019].

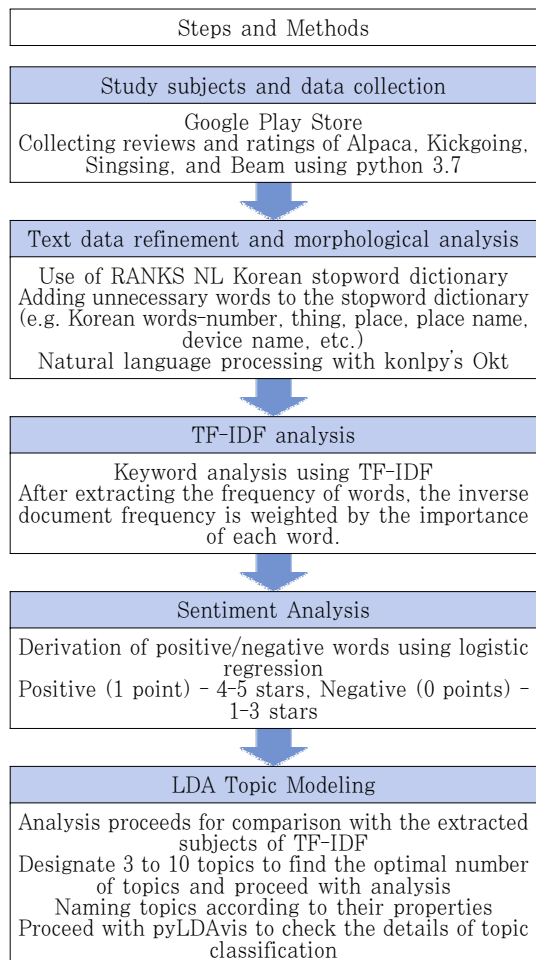
As such, research on electric scooter sharing

services is being actively conducted, but mainly studies using news articles or public data have been conducted [Kim et al., 2021]. On the other hand, this study is to analyze the success factors of the shared electric scooter service by extracting extensive review data from the reviews voluntarily written by users of the shared electric scooter service and identifying the user's intention.

3. Research Methodology

3.1 Data Collection and Preprocessing

<Table 1>



3.2 Analysis Methods

3.2.1 TF-IDF

TF-IDF (Term frequency inverse document frequency) is a technique widely used in text analysis. It is simple and has excellent performance because it can quantify the importance of words that appear frequently in a specific document and select words [Kim et al., 2018].

It is calculated as the product of the word frequency count, TF, and the inverse document frequency count, IDF, and indicates the importance of a word contained in a specific document in the entire document by digitizing it [Salton and Buckley, 1988].

$$TFIDF = TF \times \frac{1}{DF} \quad (1)$$

TF = Frequency of a specific word in a document

DF = Frequency of a specific word in multiple documents

IDF = reciprocal of DF

TF-IDF was performed using the TF-IDFTransformer package provided by the Sklearn library for the BOW (Bag of Words) that had undergone data preprocessing.

In order to identify positive and negative emotions, which are subjective elements of user reviews, the X value of the model is the evaluation of user reviews in order to predict whether the review is positive or negative, and to understand the user's emotion. It was set as the content, and the Y value was designated as the positive/negative emotion of the user. In order to divide user reviews into positive and negative, 4 to 5 points of the 5-point rating were converted into 1 point for positive evaluation. In the case of 3 points,

it is ambiguous to belong anywhere with a value between positive and negative, but in the case of review stars, 4 to 5 points account for a large proportion, so 1 to 3 points were converted to 0 points for negative evaluation [Cho et al., 2014].

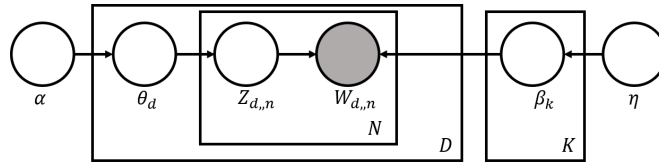
The ratio of training data and test data was distributed 7:3. As a result of evaluating the performance of the regression model using logistic learning, the accuracy of the model was calculated as 71%. Accuracy is determined by whether the predicted results match, and there is no special standard for accuracy, but in previous studies, even a new technique of 60% or more was considered worthy of research [Kang et al., 2015].

Since the results of logistic regression analysis are output as numbers, it is difficult to understand their meaning, so morphemes were extracted using index_vectorizer as the coefficient of the regression model, and 20 positive and negative words were extracted respectively.

3.2.2 LDA Topic Modeling

The LDA topic modeling technique is one of the probabilistic topic modeling techniques that indicates which topics exist in each document in a given document. Each document represents a probability distribution for words nested in a topic in order to infer which topics exist [Blei et al., 2003].

⟨Figure 1⟩ is a schematic diagram of the LDA document generation process. Gray circles represent observed variables, and white circles represent latent variables. Arrows represent relationships between nodes, and outer rectangles represent documents and represent the number of topics and words. Topic modeling using LDA is a process of finding potential topics by assuming a Dirichlet dis-



- K : Number of topics specified
- D : Number of documents
- N : Document length, Number of words
- α : Dirichlet prior weight of topic K , Parameter for determining the value of θ
- η : Dirichlet prior weight of word w by topic, β value decision value parameter
- θ_d : Percentage of topics in the d th document
- $Z_{d,n}$: The topic of the i th word in the d th document
- $W_{d,n}$: i th word in d th document (Word extracted from document)
- β_k : Generation of word w in topic k = Probability
- k : Index of each topic
- d : Index of each document 각 문서 인덱스
- i : Word index in document

<Figure 1> LDA Document Creation Process

tribution, one of the probability distributions, if there are documents and words are included in them, and putting documents and words into numbered topics one by one.

Using corpora.Dictionary() from the copra package of the gensim library, integer encoding was performed on words, and the frequency counts of words were recorded. The number of topic classifications was carried out from 3 to 10, and as a result of the progress, classification was optimized when the number of topics was 3. Topics are named according to the classified characteristics, and the numerical value of the characteristic words represents the contribution to the topic. After topic classification, pyLDAvis was run for LDA visualization. The circle on the left is the number of topics, and the distance from each circle shows the difference between the topics. Depending on the topic, characteristics appear on the right.

4. Result

4.1 TF-IDF Analysis

A TF-IDF analysis was conducted to classify the positive and negative factors of electric

scooter sharing service users. As a positive factor, it was found that commuting and short distances and public transportation were used mainly for ambiguous distances to move, and it was found that they were satisfied. It appeared that they were dissatisfied with safety issues.

<Table 2> TF-IDF Positive

	Alpaca	SSingSSing	Kickgoing	Beam
이동	1.93	최고 1.85	최고 2.54	최고 1.65
출퇴근	1.42	조금 1.79	출퇴근 1.50	확대 0.95
이용	1.4	가끔 1.52	만족 1.16	편리 0.93
수단	1.35	언덕 1.34	출근길 1.09	안정 0.9
환승	1.34	다만 1.27	거리 1.07	꿀잼 0.84
거리	1.22	만족 1.11	공유 1.00	무료 0.82
안정감	1.14	거리 1.09	출퇴근길 0.97	만족 0.77
쓰기	1.13	항상 1.03	안정 0.89	애용 0.75
택시	1.05	제일 1.00	이동 0.89	안전 0.64
안정	0.99	오르막 0.98	확장 0.86	개꿀 0.62
근거리	0.95	출퇴근 0.96	이벤트 0.86	저렴 0.6
절약	0.92	약간 0.92	재미 0.85	경험 0.59
뽀너스	0.87	편리 0.89	수단 0.83	재미 0.58
동시	0.82	가성 0.84	종종 0.82	공유 0.56
퇴근	0.82	수단 0.82	축지법 0.82	이동 0.55
버스	0.8	모터 0.79	근거리 0.81	할인 0.54
단거리	0.76	재미 0.77	단축 0.80	개편 0.5
승차	0.75	이동 0.76	출근 0.77	방법 0.5
안전	0.75	버스 0.75	강추 0.72	근거리 0.49
도감	0.75	확대 0.75	비교 0.72	합리 0.48

<Table 3> TF-IDF Negative

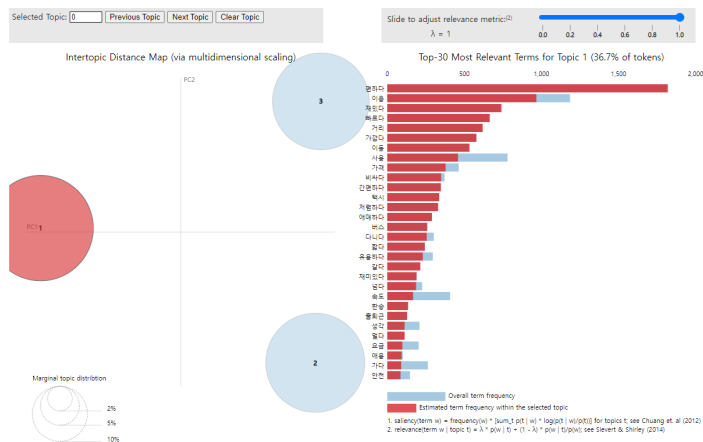
Alpaca		SSingSSing		Kickgoing		Beam	
인증	-3.15	업데이트	-3.33	연결	-3.74	불가	-2.76
최악	-2.9	인증	-2.51	인증	-3.57	금지	-2.46
방전	-2.54	최악	-2.31	가입	-2.71	최악	-2.35
업데이트	-2.54	제발	-2.15	쓰레기	-2.6	지역	-2.31
등록	-2.5	삭제	-2.07	최악	-2.58	인증	-2.24
종료	-2.3	등록	-2.03	불가	-2.56	주차	-2.18
점검	-2.26	로그인	-2.02	해제	-2.5	결제	-2.17
배터리	-2.19	탈퇴	-2	지도	-2.44	오류	-2.13
지도	-2.19	배터리	-1.98	오류	-2.36	번호	-2.12
개선	-2.07	가입	-1.96	등록	-2.22	가입	-2.05
로딩	-1.96	위치	-1.92	장난	-2.19	지도	-2.03
회원	-1.95	설치	-1.91	이탈	-2.16	어플	-2.02
지각	-1.95	운전면허	-1.87	실행	-2.13	센터	-2.02
불가	-1.88	장난	-1.81	삭제	-2.07	면허증	-1.97
해도	-1.82	제대로	-1.75	배터리	-1.98	등록	-1.89
별로	-1.77	가요	-1.66	무슨	-1.97	배터리	-1.84
고장	-1.75	면허증	-1.62	고장	-1.96	로딩	-1.84
이유	-1.75	지도	-1.59	점검	-1.96	업데이트	-1.83
구역	-1.74	반납	-1.56	회원	-1.86	종료	-1.81
입력	-1.74	코드	-1.56	업데이트	-1.79	쓰레기	-1.76

4.2 LDA Topic Modeling

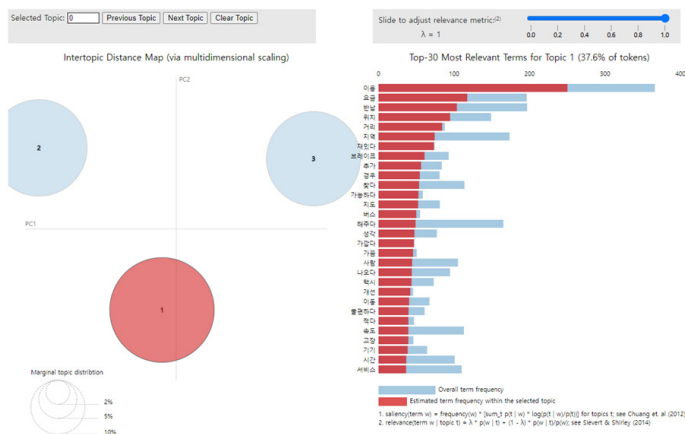
The LDA Topic Modeling analysis is a technique for extracting topics, suggesting keywords for each shared service provider. In the case of Alpaca, it was named as entertainment, safety, and service area. In the case of SSingSSing, it was named as application complaint, use complaint, and convenience. In the case of Kickgoing, it was named as application complaint, mobility, and service area. In the case of Beam, it was named as entertainment, dissatisfaction in use, and inconvenient parking.

<Table 4> 4 LDA Topic Modeling

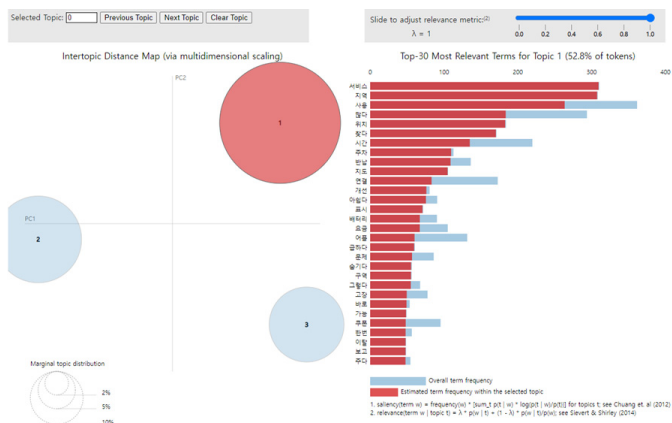
App	Topic1	Topic2	Topic3
Alpaca	오락성	안전성	서비스지역
	편하다	헬멧	지역
	0.096	0.026	0.032
	이용	안전하다	서비스
	0.051	0.014	0.024
	재밌다	수단	싸다
	0.039	0.013	0.023
	빠르다	쓰다	반납
	0.035	0.013	0.02
	거리	급하다	속도
0.032	0.011	0.015	
SSingSSing	어플불만	사용불만	편리성
	안되다	이용	편하다
	0.028	0.025	0.036
	업데이트	요금	사용
	0.022	0.012	0.022
	뜨다	반납	비싸다
	0.009	0.01	0.02
	반납	위치	가격
	0.009	0.009	0.016
	주차	거리	이용
0.008	0.008	0.012	
Kickgoing	어플불만	이동성	서비스지역
	연결	편리하다	서비스
	0.015	0.035	0.02
	인증	최고	지역
	0.013	0.034	0.02
	어플	재밌다	사용
	0.012	0.026	0.017
	많다	빠르다	많다
	0.009	0.023	0.012
	오류	비싸다	위치
0.008	0.015	0.012	
Beam	오락성	사용불만	주차불편
	좋다	편하다	주차
	0.332	0.069	0.037
	비싸다	이용	구역
	0.039	0.025	0.022
	재밌다	종료	이용
	0.034	0.021	0.017
	최고	안되다	편리하다
	0.023	0.021	0.013
	인증	자다	지역
0.008	0.013	0.012	



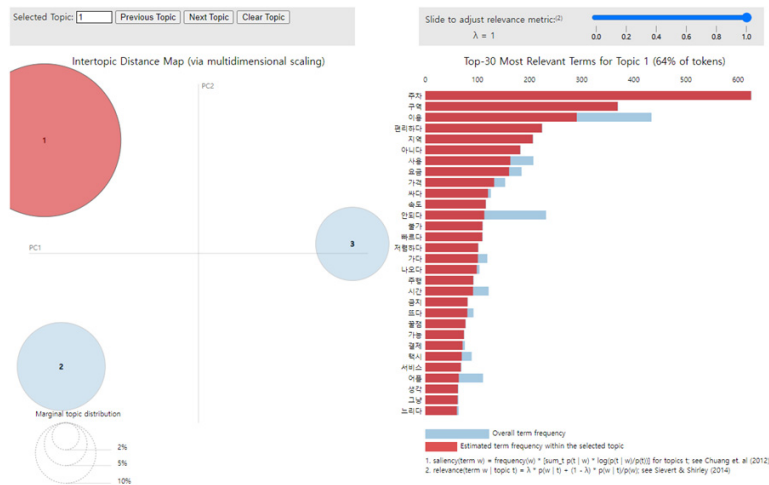
<Figure 2> Alpaca Pyldavis



<Figure 3> SSingSSing Pyldavis



<Figure 4> Kicking Pyldavis



〈Figure 5〉 Beam Pyldavis

5. Conclusion

5. 1 Conclusion and Business Implications

This study analyzed reviews, grasped users' minds, and analyzed service improvements and success factors by limiting 4 domestic companies among electric scooter service companies.

First, words that commonly appear at the top of positive words using TF-IDF are commuting, short distance, short distance, taxi, public transportation, etc. It was revealed that it is mainly used in streets where it is ambiguous to move by public transportation or taxi. Words that commonly appear at the top of negative words are authentication, update, registration, etc., and license registration is not possible due to application errors, or errors that are not authenticated often occur, and update functions due to errors are repeatedly executed. It was revealed that there was difficulty in using the device because the application could not be used.

Second, as a result of subject analysis using LAD topic modeling, alpaca was selected as

entertainment, safety, and service area. The theme was determined by expanding the use area. In the case of SingSing, the topics were set as application complaints, device use complaints, and convenience, and complaints due to application errors, GPS errors while using the device, making it difficult to find the device, and many brake errors. There is convenience in , so I decided on convenience.

As a result of each analysis with TF-IDF and LDA topic modeling, users of the electric scooter sharing service give positive evaluations that they are useful in commuting and short-distance movement and public transportation, and infinite repetition due to application errors. And rebooting, license registration error, GPS accuracy is low, it takes a long time to find the device, and device use is stopped due to battery level display error. In addition, when the length of the review is short, the accuracy of positive/negative evaluation decreases, and it can be seen that negative evaluation appears together in positive evaluation.

The implications of this study are as follows.

First, it is significant that the voluntary

reviews of customers who use the electric scooter sharing service were analyzed using TF-IDF and LDA topic modeling. Most of the existing studies are studies using questionnaires or text mining studies using public data, and text mining analysis on voluntary reviews of electric scooter sharing service users is still incomplete.

Second, it can be seen that users of the electric scooter sharing service mainly use it for quick movement during commute time, or use it as a means of transportation in a distance where it is difficult to move by public transportation or taxi. Therefore, if it is properly placed in an area where office workers commute a lot or in a section where they have to walk to use public transportation, it will be possible to secure more users.

Third, it is necessary to provide quick response and feedback to device errors and application errors. When users of the electric scooter sharing service use the application, errors in the procedure for license authentication, infinite repetition and rebooting of the application, errors in the area of use, and low accuracy of GPS take a long time to search the device, and errors in displaying the remaining battery. The phenomenon of stopping while moving appeared as a complaint. Therefore, if quick responses and feedback are provided for applications, satisfaction and reliability of service providers can be improved.

Lastly, to ensure safety, helmet introduction and safety devices should be added. In the case of alpaca, safety issues were secured because helmets were attached, but helmets were not yet introduced, so there were many requests for safety issues, and there were many reports that the impact of tires was felt by the body. Therefore, additional introduction to secure the safety of electric

scooter sharing service companies will increase the reliability of service companies.

5.2 Limitations and Future Research Directions

This study has the following limitations in interpreting the research method and research results, and intends to suggest future research directions to supplement them.

First, this study is a study using the reviews of users of the electric scooter sharing service provided on the Play Store. Among the companies using the service, 4 of the applications that were released in a similar period and had a large number of reviews were selected and reviews were collected, and analyzed. There is a limit to generality in that the analysis was limited to the top four companies. Therefore, in future research, it will be possible to make more accurate measurements by expanding the scope of companies and collecting and analyzing reviews of various sharing services.

Second, online reviews include terms related to a specific domain and various expression methods and terms used by consumers in the text. These terms appear in very diverse forms and change over time, so there is a realistic limit to reflecting all terms in the analysis. This study is also the same, and in future studies, it will be possible to build a word dictionary with domain experts through separate studies or consider artificial intelligence-based natural language processing.

Third, the topic analysis result of topic modeling may vary due to the analyst's subjective topic selection. If the analysis is performed by another analyst, the content of the topic may change due to the subjective subject selection of the analyst. Therefore, a more accurate analysis can be performed by injecting

several analysts with a lot of experience and objectivity and comparing the analysis results.

As a future research project, it is expected that various studies on techniques for extracting meaningful words will be conducted, and follow-up analyzes using other analysis models will be conducted.

References

- [1] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent dirichlet allocation", *Journal of machine Learning research*, Vol. 3, 2003, pp. 993-1022.
- [2] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., "Learning phrase representations using RNN encoder-decoder for statistical machine translation", *arXiv preprint arXiv*, 2014, 1406.1078.
- [3] Hong, D. H. and Park, K. G., "Smart mobility: The future of transportation services", *Sejong: The Korea Transport Institute*, 2011, pp. 1-244.
- [4] Kang, S. N., Kim, Y. S., and Choi, S. H., "Study on the social issue sentiment classification using text mining", *Journal of the Korean Data And Information Science Society*, Vol. 26, No. 5, 2015, pp. 1167-1173.
- [5] Kim, H. J., Lee, T. H., Ryu, S. E., and Kim, N. R., "A study on text mining methods to analyze civil complaints: Structured association analysis", *Journal of the Korea Industrial Information Systems Research*, Vol. 23, No. 3, 2018, pp. 13-24.
- [6] Kim, H. J., Song, Y. J., Lee, E. Y., Jeong, S. H., Park, E. J., Lim, H. K., "Implementation of supplement program for electric scooter rental app", *Proceedings of KIIT Conference*, 2021, pp. 455-457.
- [7] Kim, S. H., Chang, N. S., and Kim, K. W., "Academic trend analysis of shared economy based on text mining and network analysis", *Journal of the Korean Entrepreneurship Society*, Vol. 16, No. 2, 2021, pp. 15-34.
- [8] Kim, S. J., Lee, G. J., Choo, S. H., and Kim, S. H., "Study on shared e-scooter usage characteristics and influencing factors", *The Journal of The Korea Institute of Intelligent Transportation Systems*, Vol. 20, No. 1, 2021, pp. 40-53.
- [9] Kim, S. Y. and Chung, Y. M., "An experimental study on selecting association terms using text mining techniques", *Journal of the Korean Society for Information Management*, Vol. 23, No. 3, 2006, pp. 147-165.
- [10] Lee, H. S., Baek, K. H., Jung, J. H., and Kim, J. H., "User's behaviors of smart personal mobility sharing services: Empirical evidence from electric scooter sharing service", *Proceedings of the KOR-KST Conference*, 2019, pp. 462-463.
- [11] Lee, U. Y. and Kim, S. I., "A study on user experience of scooter-sharing system: Focused on kickgoing and lime", *Journal of Digital Convergence*, Vol. 19, No. 2, 2021, pp. 425-431.
- [12] Lessing, L., "Remix: Making art and commerce thrive in the hybrid economy" *Penguin*, 2008.
- [13] Park, J. M., Jeong, E. S., and Kim, J. H., "Research on the current situation and improvement of bicycle sharing platforms using big data", *Proceedings of the Korea Information and Communications Society General Academic Conference*,

- Vol. 23, No. 2, 2019, pp. 303-305.
- [14] Ryu, J. S. and Cho, W. D., "An impact analysis on shared bike utilization of COVID-19 using machine learning regression learners", Proceedings of Korean Institute of Next Generation Computing, May 2021, pp. 79-82.
- [15] Salton, G. and Buckley, C., "Term-weighting approaches in automatic text retrieval", Information Processing & Management, Vol. 24, No. 5, 1988, pp. 513-523.
- [16] Sundararajan, A., "The sharing economy: The end of employment and the rise of crowd-based capitalism", MIT press, 2017.

■ Author Profile



Kyoung-ae Seo

She is currently serving as a director at Da Vinci Ventures Co., Ltd. and is taking game QA-related classes at Dong Seoul University and Yongin Arts & Science University.

He received a master's degree in Information and Communication Engineering from Ajou University's Graduate School and a Ph.D. in business administration from Hoseo University Graduate School. Her interests are data analysis using machine learning and deep learning of big data and research on prediction using it.



Jung Seung Lee

He is currently working as an associate professor at the School of Business at Hoseo University, and serves as the president of the Korea Data Strategy Society and a director

of the Korea Intelligent Information Systems Society. He received his bachelor's and master's degrees from the Department of Management Science at KAIST, and his Ph.D. in management engineering from the same graduate school's Business School. He founded SNS site, Old Boy (oldboy.co.kr), financial consulting site, Best Money (bestmoney.co.kr), blockchain-based personal history authentication solution (Career Ledger) company, Smart-I Co., Ltd. His main interests are Smart Factory, Smart Grid and Sustainable Supply Chain.