Analysis on Review Data of Restaurants in Google Maps through Text Mining: Focusing on Sentiment Analysis

Bee Shin[1], Sohee Ryu[1], Yongjun Kim[1], Dongwhan Kim[2*]

Abstract

The importance of online reviews is prevalent as more people access goods or places online and make decisions to visit or purchase. However, such reviews are generally provided by short sentences or mere star ratings; failing to provide a general overview of customer preferences and decision factors. This study explored and broke down restaurant reviews found on Google Maps. After collecting and analyzing 5,427 reviews, we vectorized the importance of words using the TF-IDF. We used a random forest machine learning algorithm to calculate the coefficient of positivity and negativity of words used in reviews. As the result, we were able to build a dictionary of words for positive and negative sentiment using each word's coefficient. We classified words into four major evaluation categories and derived insights into sentiment in each criterion. We believe the dictionary of review words and analyzing the major evaluation categories can help prospective restaurant visitors to read between the lines on restaurant reviews found on the Web.

Key Words: Restaurant Review, Sentiment Analysis, Text Mining, Social Computing.

I. INTRODUCTION

The importance of online reviews is prevalent these days. With its ubiquity, vastness, and speed we can now perceive customer satisfaction levels at one sight without much difficulty, and even proceed to apply the feedback to improve the overall customer experience. As the influence of reviews is ever-growing, customers are greatly affected by reviews in terms of decision-making [1-2]. For example, online reviews are the primary deciding factor for customers when deciding whether or not to buy a specific product or visit a particular restaurant. We indeed encounter millions of reviews containing a complex range of assessment factors affecting the consumer's preferences in modern days. However, such information is generally provided by simple comments or mere star ratings. Based on the simple reviews, it is not easy to get an overall idea of what consumers as a whole want or what factors consumers refer to when making their revisit decisions [3].

This study proposes a methodology that breaks down comments on restaurant reviews written in Korean. We collected more than 5,000 reviews from Google Maps to perform sentiment analysis and derive four primary evaluation criteria after the analysis. These criteria include food, price, service, and atmosphere. After data cleaning, we vectorized extracted words to perform machine learning and used this trained model to calculate coefficients according to positivity and negativity. In conclusion, we built a new sentiment dictionary for each evaluation criterion.

The contribution of this study is as follows. We proposed a machine learning approach towards analyzing what are some words to look out for when searching positive or negative reviews on restaurant reviews. We provided a dictionary of such words in the tables provided. We believe the findings of this research can further assist in reading the lines between online reviews and can also be used as a tool for the marketing and strategy of restaurant managers.

II. LITERATURE REVIEW

As the importance of online reviews emerged, we were able to find and review relevant studies for our research. The study of Lee and Son [4] aimed to grasp the perception and appraisal of Seoul urban park users through text analysis. They used Google review data provided by Google Maps and determined if the reviews helped extract meaningful information for park management plans. The paper provides an excellent technical methodological approach to collect and analyze online reviews, and we also utilized similar Python programming libraries such as Selenium [5].

Selenium is an open-sourced package that supports web browser automation and is often used in crawling data from web pages.

We also reviewed studies aiming to acquire individual evaluation scores toward restaurants. So and Shin [6] informed which methods we could adopt to gather review data from Google Maps and handle Korean characters grammatically. KoNLPy, a python library to make Korean sentences split into grammatical element units, could be utilized for our research too. The literature also mentioned that its researchers used four attributes to classify reviews according to their evaluation criteria: food, price, service, atmosphere, which are initially from a travel platform TripAdvisor's review system. As people tend to evaluate restaurants within the four criteria, we decided to apply this classification system to our research. Kim et al.[7] also explored the usefulness of online reviews in the delivery application called Yogiyo and provided insights on how the research on delivery app reviews should progress, centered on fried chicken reviews.

Sentiment analysis is often used in analyzing online review data. In the study of Kuman, De, and Roy [8], tweets from microblogging sites were collected and analyzed to build a recommendation system that counts current trends, public sentiment, and user response to movies. Sentiment analysis is used to bridge the gap when a system has no previous user data to work upon. Kumar et al.[9] worked further in analyzing the result of the sentiment analysis by exploring the impact of age and gender using machine learning. Authors collected reviews on books from Facebook along with users' age groups and gender information and segmented upon the demographic information. They were able to compare the impact of different user groups in terms of expressing his/her opinions. Pathak, Kuman, Roy, and Kim [10] stepped further by conducting Aspect-Based Sentiment Analysis (ABSA) that identifies the aspects within the given sentences to the review dataset in Hindi. They used deep learning natural language models (namely BERT) on the aspect category detection task, and the results outperformed the state-of-the-art models on Hindi datasets by far. This study succeeded in empirically proving that applying machine learning can be a great way to analyze online review data.

In the study of Mathayomchan and Taecharungroj [11], we were able to take away the quantitative analysis methods used to analyze the sentiment of restaurant reviews affecting the overall customer satisfaction levels. The authors used Logistic Regression, VADER Sentiment Analysis, and Lexical Salience-Valence Analysis to measure how customers perceive the four attributes of restaurants. To complement the tests of relationships between attributes and customer experience, the research delved deeper into analyzing the underlying words in online reviews and their effects on

the overall experience through lexical salience-valence analysis (LSVA), which helps analyze words in online reviews. We retrieved restaurant selection and online reviews collection methods from the data collection and preparation stage. From the sentiment analysis stage, we used sentence extraction and attribute identification. Lastly, we utilized logistic regression from the data analysis stage to see relationships between restaurant attributes and 5-star ratings.

III. RESEARCH METHOD

The research was conducted in the following order:

Data Crawling and Preprocessing: We designed a custom data crawling algorithm using Selenium Webdriver and collected 5,427 reviews from Google Maps. We then tagged all words in the text corpus and extracted nouns with two or more letters.

Text Vectorization and Machine Learning: After creating sets of words and their frequency of appearance, we vectorized each word using TF-IDF algorithm. Vectorization allowed us to weigh words. We then applied a random forest machine learning algorithm to retrieve relatively more important words with positive and negative sentiments by calculating the coefficient value of words.

3.1. Data Crawling and Preprocessing

We crawled 5,427 reviews from Google Maps using the Webdriver function from a Python package called Selenium (Fig. 1). We made a custom Python function to extract words from the reviews: to extract words by spaces. If there was a space between words, they were considered different words. Upon applying the function to our database of 5,427 reviews, we were able to get sentences sliced into regular expressions.

Then, we extracted nouns from the regular expressions list. After that, we made a list of all nouns from all reviews and counted words to find the most frequently used ones. The result showed that words like 'taste (맛)', 'price (가격)', 'unspecified place (곳)', 'when (때)', etc. The word 'taste'

Fig. 1. Sample reviews crawled from Google Maps.

```
available_counter = Counter({x: counter[x] for x in counter if len(x) > 1})
available_counter.most_common(10)
```

```
[('가격', 705),
 ('고기', 494),
 ('신촌', 401),
 ('음식', 387),
 ('메뉴', 321),
 ('직원', 292),
 ('국물', 281),
 ('제공', 267),
 ('추천', 262),
 ('최고', 261)]
```

Fig. 2. The most common words found in reviews are longer than one letter.

showed up 1,787 times in reviews and it was the most frequent word. From the extracted nouns, we were able to find that one-letter words like '때' do not add any details in analyzing the words and decided to remove one-letter words. After the data cleaning, the most frequent words were 'price (가격)', 'meat (고기)', 'Sinchon(신촌)', 'food (음식)', 'menu (메뉴)', and 'waiter and waitress (직원)' (Fig. 2).

3.2. Text Vectorization and Machine Learning

We then vectorized the extracted words using the words list and their counts. We made a dictionary named word_count_list, which paired up the extracted words to the number of times they were used (Fig. 3).

For a more thorough analysis of word frequency analysis, we applied TF-IDF vectorization. TF-IDF is often used in finding relatively more important words in the collection of documents. TF-IDF measures Term Frequency through two methods: Term Frequency and Inverse Document Frequency [12]. Term Frequency measures how often a word w occurs in a document d, and Inverse Document Frequency measures the importance of the word. Put together, TF-IDF can weigh and value the importance of a word in a corpus, and therefore, it helps to find relatively more important words in all reviews.

Then, for machine learning, we labeled all comments to 0 or 1. We assigned 1 to comments that have ratings bigger than 3. This means comments with ratings 4 and 5 are labeled with 1. These comments will be a collection of positive comments. The comments with ratings 1~3 were labeled 0 as comments with negative connotations. As a result, we ended up with 1,257 negative comments and 4,170 positive comments.

We then split our data for the machine to learn by making

```
word_count_dict = dict(zip(word_list, count_list))
word_count_dict
```

```
{'가가': 3,
 '가게': 166,
 '가겐': 1,
 '가격': 706,
 '가계': 1,
```

Fig. 3. Sample word in the word_count dictionary.
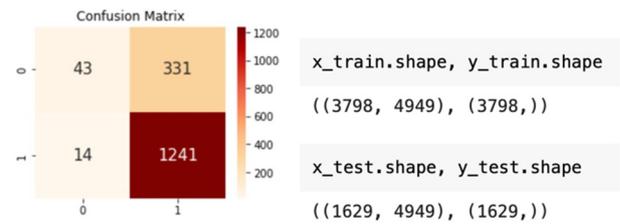
Fig. 4. Confusion matrix for machine learning.
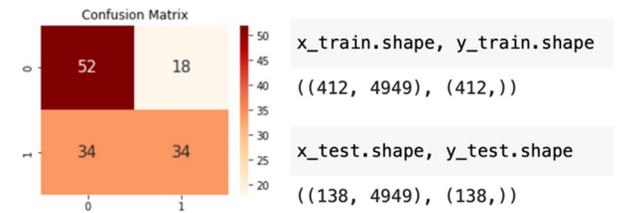
Fig. 5. Confusion matrix for machine learning after imbalance adjustment.

a test dataset with 30% of the total data. Fig. 4 is the confusion matrix generated after machine learning.

Through this, we were able to find out that the model tends to predict comments excessively positively (the lower right corner of the confusion matrix). In other words, positive reviews are well predicted, but the prediction accuracy for negative reviews was very low. We found that the imbalance in sample data size creates this skewness. We randomly sampled 275 of each positive and negative data to adjust the imbalance. This would help the accuracy of machine learning since we now have a balanced sample size for positive and negative reviews. After adjusting the imbalance, we applied machine learning again and found that we now have a similar prediction rate for both positive and negative data (Fig. 5).

After returning the index number and elements of the collection in a tuple form, we extracted coefficients according to their positivity and negativity. We retrieved a list of 100 relevant positive and negative words in descending order. We then used this trained algorithm to analyze the sentiment of 5,427 reviews collected in the database.

IV. RESULTS

As a result of our research, we classified the restaurant reviews in four major categories using the classification method used in [So, J. S., & Shin, P. S.]. The categories were food, price, service, and atmosphere. To illustrate, for example, along with food menus, words such as '*slang* very tasty (핵존맛)', 'tasty place (맛집)', 'generously (듬뿍)' were classified into the 'food' category. And for price, 'comparing to (대비)', 'cost-effective (가성비)', 'reasonable

(합리)' were sorted into this category. Next, 'all-you-can-eat (무한리필)', 'waiter/waitress (직원)', 'quality (품질)' were sorted into the 'service' category, and finally 'warmth (정감)', 'atmosphere (환경)', 'mood (분위기)' were classified into the 'atmosphere' part. There were also neutral words such as 'best (최고)', 'real (진짜)' and these terms could belong to any one of the four categories. We added these words to a dictionary of positive neutral words.

4.1. Positive Sentiment Words

Table 1 is the positive sentiment words that we selected that provide some information to distinguish positive nuance.

Table 2 is the collection of other classifications and words: price, service, and atmosphere.

4.2. Negative Sentiment Words

Tables 3 and Table 4 are the classification of negative-sentiment words into four categories as we did with positive words. Words representing the menu of restaurants were categorized into food category, and words also like 'size (사이즈)', 'texture of food (식감)', 'tough (질김)' were classified (Table 3). In terms of price section, there were negative terms like 'cost-inefficient (창렬)', 'cost-effective (가성비)'. Also, 'unkind (불친절)', 'serving (서빙)', 'facial expression (표정)' were classified into the service category, and lastly 'sanitation (위생)', 'small (협소)' were categorized into the

Table 1. Classification of positive sentiment words – Food.

Food	
Cheese (치즈)	Tasty place (맛집)
BBQ (구이)	Fish (물고기)
Spicy (매콤)	Beer (맥주)
Squid (오징어)	Soup (수프)
Mackerel (고등어)	Rice noodle (쌀국수)
Cream (크림)	Quality (품질)
Rice with pork (부타동)	Brazil (브라질)
Slang very tasty (핵존맛)	Stir-fried vegetables (잡채)
Gimbap (김밥)	Noddle with meat (고기국수)
Generously (듬뿍)	Port cutlet (돈까스)
Sundae (순대)	Spanish Mackerel (삼치)
A type of sundae (피순대)	Soft tofu stew (순두부찌개)
Snow crab (대게)	Galbi soup (갈비탕)
Crunchy (바삭)	King crab (킹크랩)
Braised galbi (갈비찜)	Steak (스테이크)
Slang very tasty (개졸맛)	Bulgogi (불고기)

Table 2. Other classifications of positive sentiment words.

Classification	Word
Price	Price (가격)
	Comparing to (대비)
	Reasonable (합리)
	Cost-effective (가성비)
	11,000 won (만천원)
Service	Waiter/waitress (직원)
	All-you-can-eat (무한리필)
	Sex ratio (성비)
	Quality (품질)
	Nationality (국적)
Atmosphere	Sinchon (신촌)
	Possible (가능)
	Warmth (정감)
	Tradition (전통)
	Exemplary (표본)
	Visit (방문)
	Time (시간)
	Alley (골목)
	Family (가족)
	Atmosphere (환경)
	Jeju island (제주도)
	Mood (분위기)

Table 3. Classification of negative sentiment words – Food.

Food	
Food (음식)	Menu (메뉴)
Kkakdugi (깍두기)	Red soup (홍탕)
Even the quantity is … (양도)	Small-restaurant (밥집)
Sirloin (등심)	Chuncheon (춘천)
Michelin (미쉐린)	Makguksu (막국수)
Beef (소고기)	Size (사이즈)
Though, the quantity is …(양은)	Soy sauce (간장)
Place (장소)	Bean-bulgogi (콩불)
Tough (질김)	Mixed (비빔)
Flour (밀가루)	Duruchigi (두루)
Hot spicy meat stew (육개장)	Texture of food (식감)
Vegetable (야채)	Rice ball (주먹밥)
Sweet potato (고구마)	Spicy sauce (다대기)
Leaves of cabbage (우거지)	Can (통조림)
Very small amount (쥐꼬리)	Frozen (냉동)

Table 4. Other classifications of negative sentiment words.

Classification	Word
Price	Cost (비용)
	Cost-inefficient (창렬)
	Cost-effective (가성비)
	Cheap (저렴)
Service	Unkind (불친절)
	Rotation (회전)
	Attitude problem (불량)
	Serving (서빙)
	Kindness (친절)
	Facial expression (표정)
	Waitress (아주머니)
	Operation (운영)
	Opponent (상대)
Atmosphere	Transfer (이적)
	Subway station (지하철역)
	Near (근처)
	Sanitation (위생)
	Franchise (체인점)
	Basement (지하)
	Sign board (간판)
	Far (멀리)
	Toilet (화장실)
	Concept (컨셉)
	Foreign (외국인)
	Small (협소)
	Stairs (계단)
	Mind (정신)

atmosphere part (Table 4). Like positive words, there were neutral negative words that seemed to correspond to more than two categories. For instance, 'not really (별로)' and 'not very (그다지)' could fall under all four categories, and removed from the tables.

4.3. Critical Insights for Each Review Category

At the end of our overall research process and looking at the results of 100 positive and negative words, respectively, there were some critical insights for each category that we took from them, as Fig. 6 shows.

First of all, in terms of food, we found out most words represent the food menu such as 'rice noodle (쌀국수)' and 'rice with pork (부타동)'. As it is natural to include the food name in reviews, it seemed that there were many words resulting in the food category. Also, people, especially the Korean language use various prefixes for the same root of the word. For instance, people put '핵' and '개' at the front of the word 'tasty' which are slangs for 'very tasty'. Other than food menus, we noticed that customers significantly regard the taste and amount of food.

There were not many words to describe price-related comments in the price category. However, we could see how customers thought 'cost-effective (가성비)' to be a critical factor rather than the price of food itself.

In terms of service, there were more words in negative ones than positive ones. Therefore, we thought people were likely to complain or tackle more when the service was terrible, whereas they rarely referred to the service aspect when the service was good. Finally, for the atmosphere section, there were takeaways in that customers put stress on 'accessibility', 'sanitation', and 'mood' of restaurants, the key elements that restaurants should pay attention to. On top of that, we observed that unclean toilet status and the resemblance of local restaurants greatly influenced customers' experience.

In terms of positive and negative neutral words, we could
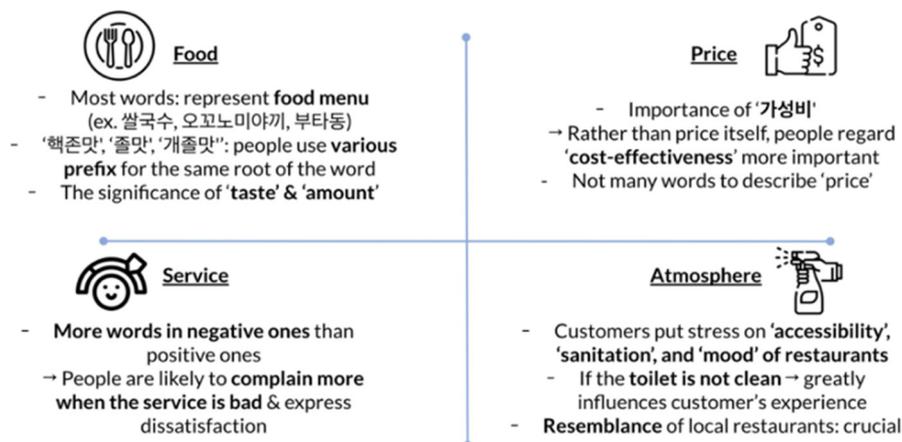
Fig. 6. Insights found for each category.

Table 5. Positive neutral words list.

Food	
Best (최고)	Most (가장)
Past (옛날)	For real (정말)
Real (진짜)	Very (아주)
Deduction (감점)	That (그것)
Restaurant (식당)	Directly (직접)
Once (한번)	Belief (소신)
Quite true (명불허전)	Mostly (대체로)
Often (자꾸)	Weekend (주말)
Each (각각)	Always (항상)
Very (무지)	Complement (칭찬)
Graduation (졸업)	Restaurant business (요식업)

Table 6. Negative neutral words list.

Food	
Thought (생각)	Little (조금)
Not really (별로)	Just (그냥)
Ordinary (보통)	Memory (기억)
Suitable (적당)	Friend (친구)
Normal (평범)	Not very (그다지)
University (대학교)	Lunch time (점심시간)
Self (자체)	Solution (해결)
Use (이용)	Case (경우)
Portion (부분)	Worst (최악)
General (일반)	Sense (느낌)
Can't possibly (도저히)	Rumor (소문)

witness apparent differences between the two groups, such as nuance and connotation. Moreover, within favorable terms, people tended to use adverbs and adjectives to express their satisfaction in a straightforward manner (Table 5). Reviewers tried to emphasize or exaggerate their positive feelings by adding adnominal phrases.

On the other hand, within negative terms, people represented their emotions in a relatively neutral way, rather than outspeaking negatively in a direct way (Table 6). Namely, they used words like 'just', 'normal' that have neutral annotation instead of 'horrible (끔찍)', 'severe (지독)' that have entirely negative nuance.

V. CONCLUSION

5.1. Implications

In the final stage of our research, we came up with a few

implications our research holds, which can be divided into two big categories.

The first implication is that machine learning could determine which kinds of vocabularies have either positive or negative sentiments. This can later be helpful to evaluate any other natural language review data from other restaurants, even when there is no rating system. Community services or bulletin-board style services could also utilize the sentiment dictionary we propose to further develop a review or ranking system. Furthermore, this machine learning skill can be applied to restaurant reviews and too many different types of analysis on data from various fields in human society.

Moreover, the second implication is recognizing what factors people usually focus on when they evaluate a restaurant. For example, the number of negative expressions that describe the service of a restaurant was more prominent than that of positive expressions. We could anticipate that people react more sensitively to dissatisfaction from bad service than satisfaction from good service. This kind of insight about customers' evaluation criteria will be necessary, especially a restaurant's marketing or business strategy.

5.2. Limitations and Conclusion

In terms of limitations, the main limitation will be precision. This problem is derived from people's subjective rating scores. For example, some people give 4 points out of 5 even when they left quite negative reviews, and similarly, some give 3 points even when they left quite positive reviews. In this research, we just regarded point 4 or over as a positive rating and three or less as negative. As not every person had the same point boundary when they judged and gave rating scores, a few vocabularies that had a quite negative nuance were regarded as positive, and the opposite also happened.

Secondly, in the process where we classified which words have which attribute of evaluation criteria (food, price, service, and atmosphere), we sorted positive and negative words into four attributes by the subjective decisions among researchers. A further study should be conducted to collect and analyze big data on restaurant reviews for more detailed and targeted classification and evaluation methods.

Lastly, we evaluated the importance of review words with coefficients calculated using a random forest algorithm. We did not go deeper in comparing the accuracy of different machine learning algorithms since much research indicated that it performs better than other common algorithms such as logical regression [13]. However, further research should be conducted to find the most appropriate algorithm for review data analysis.

In this study, we explored and broke down the real sentiment below restaurant reviews found on Google Maps. After collecting and analyzing 5,427 reviews, we vectorized

the importance of words and applied machine learning to find the coefficient of positivity and negativity of words used in reviews. We also classified words into four major evaluation categories in order to classify words that are not directly related to food. As the result, we provided what are the words that help to sense the positivity and negativity towards evaluating restaurants in terms of food, price, service, and atmosphere.

ACKNOWLEDGEMENT

REFERENCES

[1] W. Duan, B. Gu, and A. B. Whinston, "Do online reviews matter?-An empirical investigation of panel data," *Decision Support Systems*, vol. 45, no. 4, pp. 1007-1016, 2008.

[2] S. J. Heo and H. J. Bae, "Analysis of the consumption pattern of delivery food according to food-related lifestyle," *Journal of Nutrition and Health,* vol. 53, no. 5, pp. 547-561, 2020.

[3] J. H. Jung, H. I. Chung, and Z. K. Lee, "An analysis of mobile food delivery app "Baemin" by using text mining and ARIMA model*," Journal of Digital Contents Society*, vol. 22, no. 2, pp. 291-299, 2021.

[4] J. K. Lee and Y. H. Son, "Perception and appraisal of urban park users using text mining of Google maps review - Cases of Seoul Forest, Boramae Park, Olympic Park," *Journal of the Korean Institute of Landscape Architecture*, vol. 49, no. 4, pp. 15-29, 2021.

[5] Selenium, https://www.selenium.dev/, 2022.

[6] J. S. So and P. S. Shin, "Rating prediction by evaluation item through sentiment analysis of restaurant review," *Journal of the Korea Society of Computer and Information*, vol. 25, no. 6, pp. 81-89, 2020.

[7] J. G. Kim, E. C. Choi, S. H. Yoon, Y. B. Lee, and D. W. Kim, "Behind chicken ratings: An exploratory analysis of Yogiyo reviews through text mining*," Journal of the Korea Contents Association*, vol. 21, no. 11, 2021.

[8] S. Kumar, K. De, and P. P. Roy, "Movie recommendation system using sentiment analysis from microblogging data," in *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 915-923, 2020.

[9] S. Kumar, M. Gahalawat, P. P. Roy, D. P. Dogra, and B. G. Kim, "Exploring impact of age and gender on sentiment analysis using machine learning," *Electronics*, vol. 9, no. 374, 2020.

[10] A. Pathak, S. Kumar, P. P. Roy, and B. G. Kim, "Aspect-based sentiment analysis in Hindi language by ensembling pre-trained mBERT models," *Electronics*, vol. 10, no. 2641, 2021.

[11] B. Mathayomchan and V. Taecharungroj, "How was your meal?" Examining customer experience using Google maps reviews," *International Journal of Hospitality Management*, vol. 90, 2020.

[12] A. Rajaraman and J. D. Ullman, "Mining of Massive Datasets," Cambridge University Press, 2011.

[13] R. Couronné, P. Probst, and A. L. Boulesteix, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1-14, 2018.

AUTHORS

Bee Shin is a bachelor's student majoring in Creative Technology Management at Yonsei University.

Her research interests include NLP with a focus on sentiment analysis and computational public policy.

Yongjun Kim is an undergraduate student in the Creative Technology Department at Yonsei University.

His research interests include data science and social computing.

Sohee Ryu is a candidate for a bachelor's degree from the Creative Technology Department, Yonsei University.

Her research interests include big data analysis techniques and data-based solution suggestions.

Dongwhan Kim is an assistant professor in the Graduate School of Communication and Arts at Yonsei University. He obtained his Ph.D. in Communication from Seoul National University in 2017.

His research area includes Human-Computer Interaction, social computing, computational journalism, information visualization, and UX/service design.