[2023 Fall semester Graduation research] KAIST Graduation research

Graduation Research Project Final Report

Creative Project

GPT-생성 리뷰를 이용한 텍스트 기반 접근을 통한 지도 앱의 레스토랑 리뷰의 가짜 리뷰 탐지와 분석.

Fake Review Detection and Analysis of Restaurant Reviews in Map Using a Text-Based Approach with GPT-Generated Reviews.

Contents

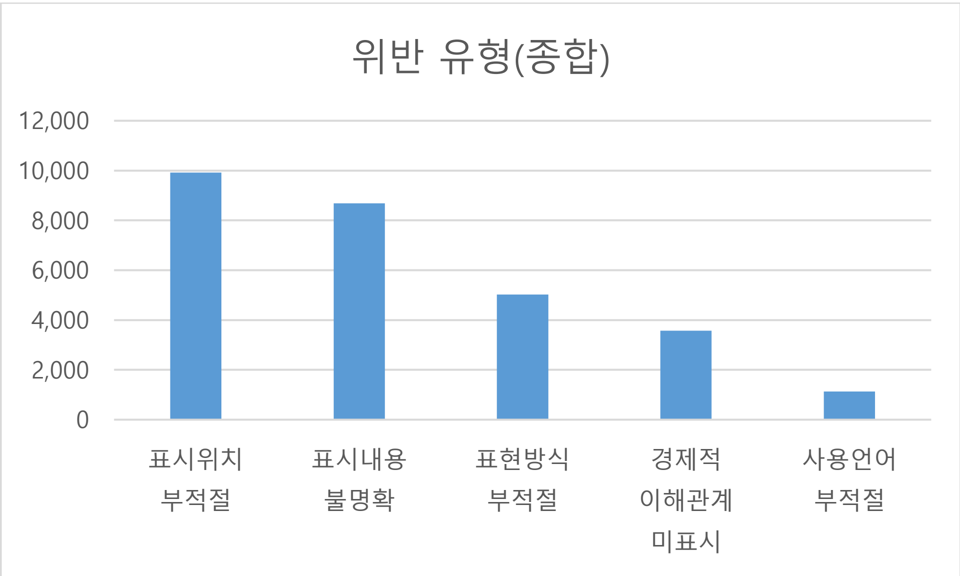
Abstract

Due to increasing advertisements in various reviews that make user experiences retreat, many researchers are trying to figure out how to classify real and fake reviews. Many researches have been done about English reviews and there are good results that is well represented by Yelp dataset. However, there are not enough researches about Korean reviews. This is mainly because of two reasons: Lower performance of Korean GPT model compared with English GPT model and lack of labeled review dataset to rely on, which can be easily found in English. Here, we used real reviews with fake reviews generated by GPT 3.5. With these, we used GPT models (KR-SBERT-V40K-klueNLI-augSTS, ko-sroberta-multitask, and paraphrase-multilingual-MiniLM-L12-v2) to get sentence similarities of reviews and then classified them based on K-medoids algorithm. Based on this model, this research made a test review dataset by 5 experiment participants and classified it. As a result, around 75% of test reviews were classified as real reviews. In conclusion, there were some similarities between GPT-generated reviews and real reviews that can be used to find some fake reviews but it isn't good enough to figure out most of them.

1. Research Background

It is well proven that online customer reviews have a positive and significant impact on customers' dicisions.[1] This has created a huge market for review advertising, and consumers are starting to be tricked and harmed by the fake reviews that are being created. At first, consumers were doing things like sharing which reviews are fake if they use certain sentence or words, or whose reviews are trustworthy. However that was only limited to individuals. Finally a few years ago, fake reviews have been started to constrain it institutionally and systematically.

Firstly, as of 1 September 2020, the Fairtrade Commission has implemented the new "Amendments to the Guidelines on the Disclosure of Recommendations, Endorsements, and Examination of Advertisements", which require content creators to disclose any financial interests in products they feature.[2] This is why, on NAVER Blog and Instagram, which are the main targets of the crackdown, if a review is sponsored by a company it must be disclosed in a prominent place. However, in February this year, the Fairtrade Commission announced the results of regular inspection of unfair advertising on social media, and it was found that many companies were not complying with this rule.[3] Between April and December 2022, the Fairtrade Commission detected a total of 21,037 illegal advertising posts, of which 9,445 were detected on NAVER Blogs, indicating that many advertisements are conducted on NAVER platforms. **[Figure 1]**



[Figure 1] Number of fake advertising violation types detected

Secondly, NAVER, which provides the Maps app, changed the review process to require a receipt or payment history to create a review from October 2021, and abolished the 1-5 rating system in favor of allowing users to choose keywords for their reviews, reducing the need for advertisers to manage their ratings. But even so, fake reviews still cause problems. It's being managed as a marketing tool, offering to create fake receipts and send them to advertising companies who will then create fake reviews for 5,000 won per review, or rewarding consumers for writing good reviews and checking them.[4]

Despite this growing consumer awareness of fake reviews and the legal and systemic improvements being made to address them, there are still many issues that need to be addressed. NAVER has also created a review cleansing system to detect patterns such as fake visit verification and advertising reviews and remove related reviews, but there are still many problems.

When looking at the areas of unfair advertising in the Fairtrade Commission's press release in **[Figure 1]**, the most common types of financial interests were "inappropriate display location", "unclear display content", "inappropriate expression method", and "non-disclosure". If it's marked, it's easy to improve by warning them to fix it, but if it's 'unmarked', it's incredibly difficult to determine if they have a financial interest, so this research will be working on that. The most common areas of misleading advertising are cosmetics and dietary supplements in goods, and food services such as restaurants in services.[4] In addition, advertisers write reviews for both NAVER Blog and NAVER Maps under the name of NAVER Place Review Management, but currently, unfair advertising management is only done for NAVER Blog posts. So, if an advertising company writes a review on NAVER Blog and NAVER Maps, the NAVER Blog post will have an economic interest disclosure, but the NAVER Maps review will not have an economic interest disclosure. Therefore, we will conduct a study to classify fake restaurant reviews on NAVER Maps, which are highly vulnerable to fraudulent advertising.

Before we can categories fake reviews, we need to define what a fake review is. The definition of fake reviews, which are most prevalent in Korea, is "reviews that do not disclose that the reviewer has received a benefit in exchange for reviewing a particular product, regardless of whether the review is truthful or not," as defined by another study.[5] And one of the most vulnerable places to these fake reviews is restaurant reviews on NAVER Maps, which is the subject of this study.

2. Research Purpose

In this study, we analyzed restaurant review data from NAVER Maps and fake reviews generated by GPT. We obtained reviews from NAVER Maps through crawling and generated fake reviews through Chat GPT 3.5 based on Yelp data from overseas. The sentence similarity of these real and fake reviews was obtained using several Korean GPT models. Based on this, we wanted to cluster them using the K-medoids learning method to find out which ones tend to be suspected fake reviews.

3. Methodology

A. Acquire NAVER Maps Review Data

For our experiment, we used reviews from Naver Maps. To ensure that our data contained a certain level of fake reviews, we used the following criteria to obtain reviews. We selected 10 restaurant with more than 1,000 reviews on Naver Maps, where at least 3 out of 5 recent blog reviews of the restaurant indicated a financial interest. We collected 14,537 reviews, as shown in **[Table S1]**, which were subsequently used in the crawl operation, and then removed reviews of 20 characters or less, which would not reveal textual features well, to obtain a total of 9,234 reviews.

B. GPT Based Fake Review Generation

Initially, we tried to generate reviews based on Korean GPT, but when we tried to generate delivery restaurant reviews based on KoGPT-2, which is the latest model among public Korean libraries, we encountered an overfitting problem.[5] For this reason, we used the YelpNYC dataset to train GPT with an English language base approach. The YelpNYC dataset was first used by Rayana and Akoglu to obtain 359,052 reviews for 923 restaurants in New York City, and the anti-fraud filters applied by Yelp to it are rated as near-ground-truth.[7] Based on this dataset, we filtered out reviews with similar conditions to the fake reviews we are targeting in this study, which are reviews that were labelled as fake but gave the restaurant a good rating. We then fine-tuned the GPT-3.5-turbo-1106 model using 16,025 reviews as additional training data, and asked it to generate reviews in English for 10 restaurants. For each restaurant, we provided information about what kind of restaurant it is and what kind of menu it serves, generating 9,234 data points, the same number as the processed data in **[Table S1]**. After that, the English data was first translated using a DEEPL translator and secondly manually corrected for awkwardness. The main changes we made were to make sure proper names, such as restaurant names, translated well, and words that are specific to the New York Review: New Yorker, New York, Manhattan, etc. have been replaced with appropriate words such as Korean, Seoul, Gangnam, etc.

.

C. Sentence Similarity Measurement and K-Medoids Clustering

A total of 18,468 data were obtained and their similarity was measured using three language models: kr-SBERT-V40K-klueNLI-augSTS, ko-sroberta-multitask, and paraphrase-multilingual-MiniLM-L12-v2.[8~10] The previous two models are Korean-only models, both of which are at the top end of the public libraries with scores around 85 on the KorSTS benchmark[11], a performance evaluation of sentence similarity measures.

The last model is the multilingual model used in many studies. After vectorising the reviews through these three language models, we measured the cosine similarity between each review.

Based on this cosine similarity, we performed K-medoids clustering, which is a clustering technique based on distance from each other. We performed fine tuning based on the Faster PAM[12] method and determined that a cluster was a cluster of fake reviews if more than half of the clusters were fake reviews. In this way, based on the Loss and Recall values for fake reviews, we obtained five models with good results and used them for the next test.

D. Testing with Real Written Reviews

To see the real-world results of the model we created above, we asked five men in their 20s to write a review. The five men visited the same two restaurants and wrote two reviews each, creating a total of 20 ground-truth reviews. We then fed this review into the models we created earlier to see how they would judge it, and for each language model, we selected the one that gave the best results.

4. Results & Discussion

A. Performance of Models

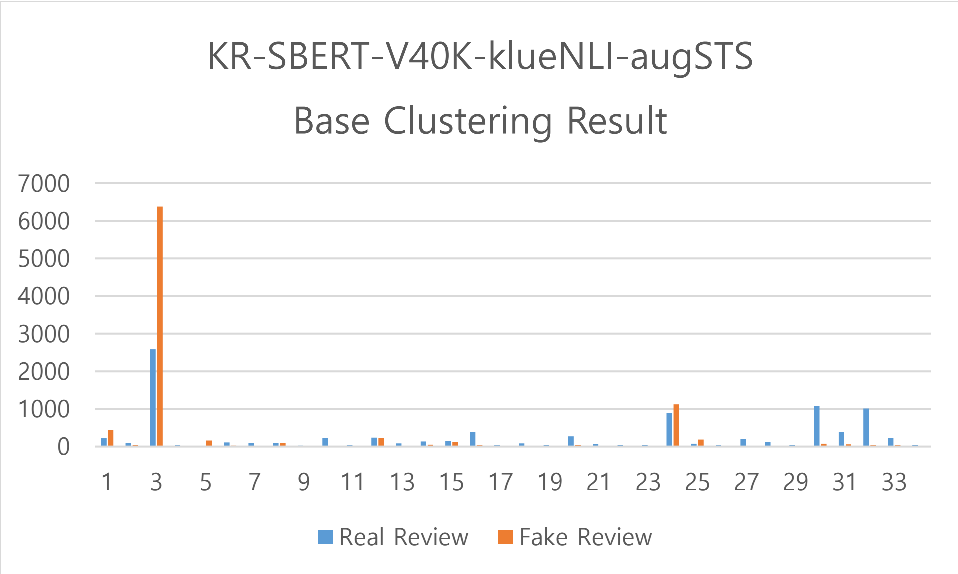
Due to the nature of the data with a mix of positive and unknown labels, there were three metrics that could be interpreted as significant in the classification performance metrics. 1. The value of Loss Function after cluster creation, 2. The recall of GPT-generated fake reviews, and 3. The recall of test reviews. So I tinkered with the number of clusters and the initial state of the model to find the state that produced the best results, as shown in **[Table S2]**.

All three models performed best in clusters of around 35. We also found that GPT-generated fake reviews performed very well at finding fake reviews, with recall values of 0.898, 0.871, and 0.931, respectively. When we tested the clustered models based on the test reviews, the models achieved recall of 0.8, 0.75, and 0.7, respectively. Finally, the models predicted 0.410, 0.365, and 0.366 of the actual collected reviews as fake. Based on these results, the similarity calculation based on kr-SBERT-V40K-klueNLI-augSTS is the most effective at catching fake reviews among the three models.

B. Characteristics of Cluster

a. KR-SBERT-V40-klueNLI-augSTS Base Model

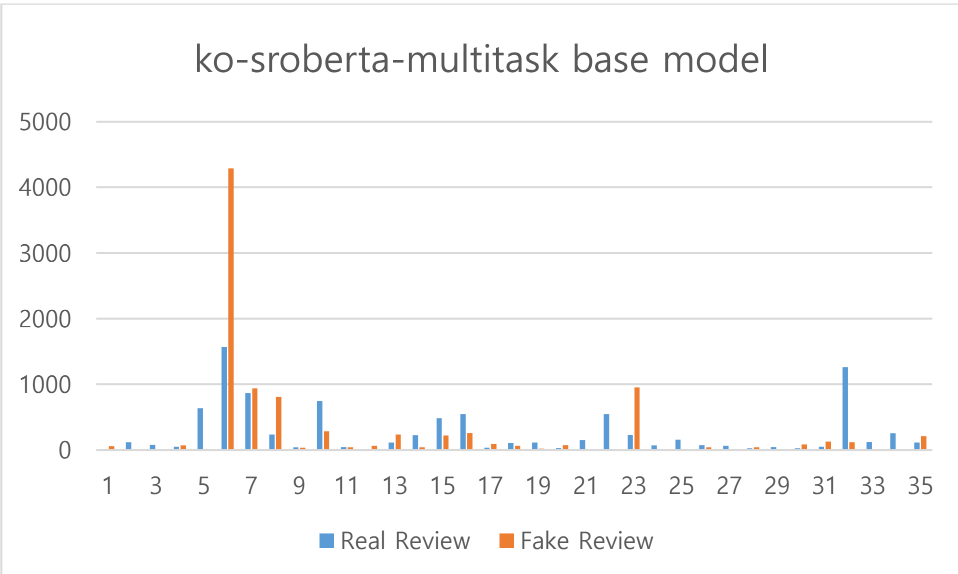
[Figure 2] is the result of clustering using this language model. Among the clusters judged to be fake, the clusters containing the most fake reviews are 3, 24, and 1, and among the clusters judged to be real, the clusters containing the most real reviews are 30, 32, and 31. Of the test reviews, the four reviews that were determined to be fake were classified as one in cluster 3, two in cluster 24, and one in cluster 25.



[Figure 2] Number of Real and Fake Reviews in Clusters Based on KR-SBERT-V40-klueNLI-augSTS Model

b. Ko-sroberta-multitask Base Model

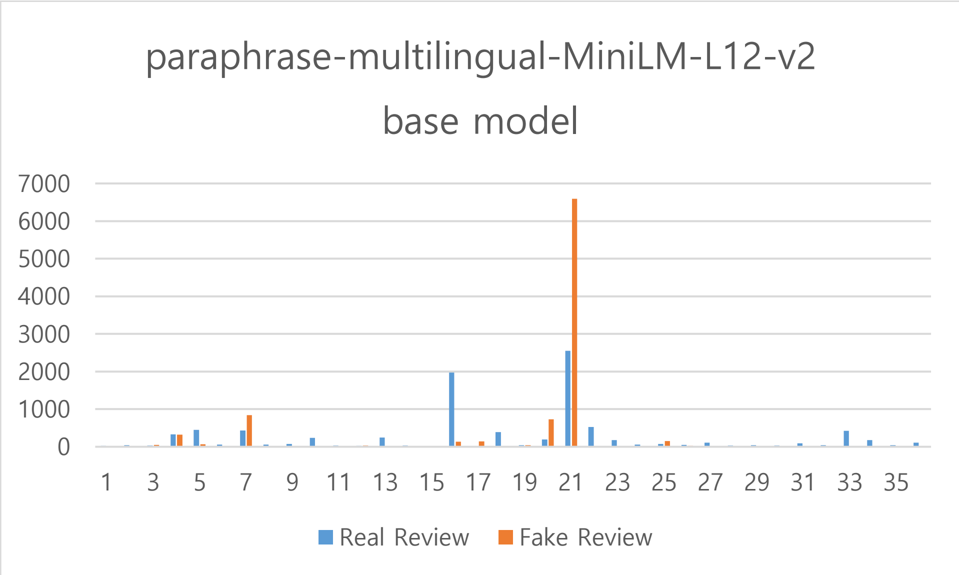
[Figure 3] is the result of clustering using this language model. Among the clusters judged to be fake, the clusters with the most fake reviews are 6, 23, and 7, and among the clusters judged to be real, the clusters with the most real reviews are 32, 5, and 16. The five test reviews that were determined to be fake were classified as four in cluster 6 and one in cluster 17.



[Figure 3] Number of Real and Fake Reviews in Clusters Based on ko-sroberta-multitask Model

c. **Paraphrase-multilingual-MiniLM-L12-v2 Base Model**

[Figure 4] is the result of clustering using this language model. Among the clusters judged to be fake, the clusters containing the most fake reviews are 21, 7, and 20, and among the clusters judged to be real, the clusters containing the most real reviews are 16, 5, and 3. The six test reviews that were determined to be fake were classified as four in cluster 21 and two in cluster 7.



[Figure 4] Number of Real and Fake Reviews in Clusters Based on Paraphrase-multilingual-MiniLM-L12-v2 Model

C. Characteristics of Clusters

We analyzed the results of the KR-SBERT-V40K-klueNLI-augSTS model, which was the best among the three models we used earlier. We used the stratified random sampling method to check the features of the clusters classified by K-medoids. 641 samples are required to represent 18468 data with 99% confidence and 5% error margin. When selecting representative samples for K-medoids clustering, it can be seen that the clusters are better represented if the medoids of the clusters are included than if they are all randomized.[13] Therefore, when selecting representative values from each cluster, select a number proportional to the size of the cluster, but make sure to include the medoids.

We used Clova Sentiment provided by NAVER to check the sentiment levels of positive, negative, and neutral sentences, and to see how the sentiment of sentences changes within a review. As shown in **[Figure 5]**, you can see the overall sentiment of the review and how positive, negative, or neutral each sentence is. In this case, we see a 50/50 mix of positive and negative sentiments, with a negative sentence followed by a positive one.

{"document":{"sentiment":"positive","confidence":{"negative":49.999733,"positive":50.00009,"neutral":1.7423878E-4}},"sentences":[{"content":"조용조용한 분위기는 아니에요.","offset":0,"length":16,"sentiment":"negative","confidence":{"negative":0.99996793,"positive":1.1239794E-5,"neutral":2.0908654E-5},"highlights":[{"offset":6,"length":9}]},{"content":" 하지만 인테리어도 예쁘고 음식도 맛있고 가성비가 좋고 예쁜 꽃도 주시고 전반적으로 만족합니다!","offset":16,"length":53,"sentiment":"positive","confidence":{"negative":0.0,"positive":1.0,"neutral":0.0},"highlights":[{"offset":15,"length":25}]}]}

[Figure 5] Example of CLOVA sentiment analysis.

We averaged the negative, positive, and neutral values of all the reviews in the sampled data by cluster, and calculated the flow of the reviews (positive sentences followed by negative sentences, positive sentences followed by neutral sentences, and so on). Among them, we found four clusters, numbered 3, 24, 27, and 30, which had prominent features in the test reviews. Clusters 3 and 24 are classified as fake reviews, and clusters 27 and 30 are classified as real reviews. Adding cluster 10, which has a characteristic appearance, we can see the average sentiment values of the five clusters in **[Table S4]**, and the sentiment flows in **[Table S5]** through **[Table S9]**, where the values 1-9 represent positive-positive, positive-neutral, positive-negative, neutral-positive, neutral-neutral, neutral-negative, negative-positive, negative-neutral, and negative-negative.

We can see that the reviews generated by GPT have a very strong positive sentiment, as they are mostly positive. In most cases, the clusters with the most fake reviews have a positive sentiment of over 90 points. We can also see that the sentence flow is also extreme, with positive-positive, positive-neutral, and neutral-positive over 90% of the time for the fake clusters, while the sentiment is much more varied for the genuine clusters.

We further examined each of these clusters and found that some of the clusters that were classified as genuine had similar characteristics to the clusters that were classified as very fake. As we saw above, we can expect to see clusters of fake reviews that are very positive and skewed toward positive-positive, positive-neutral, and neutral-positive sentiment. However, we still see reviews that fall into a gray area. Cluster 10, for example, is a cluster that leans heavily to one side, with 230 real reviews and 17 fake reviews. However, as you can see in **[Table S4]**, the positive values are so high that they look like fake reviews, but they actually have a wide range of sentiment as you can see in **[Table S9]**, which is similar to the values in **[Table S8]** for the real cluster, cluster 30. In the case of cluster 10, there were quite a few reviews in the sampled reviews that could be judged as genuine by human, such as *"직원들도 친절하신편이고 잠발라야 살짝 매콤하면서 새우랑 소시지가 많이 들어가 있고 반숙 터트리니깐 존맛"*, *"인테리어,분위기 좋고요(야외 테이블도 있어서 좋아 보였어요) 전반적으로 맛있는 편이고 양 많고요. 특별함이 있진 않지만 나쁘지 않습니다. 개인적으로 별로였던 점은 메뉴 선불로 주문하는 건데*

홀직원과 주방직원 미리 소통은 안하시는 지 이미 주문한 메뉴 안된다하여 취소 후 재 다른 메뉴 재 결재한 것/ 재결재한 메뉴에 감자튀김이 엄청나게 나오는 지 몰랐는데 저희가 이미 감자튀김 올라간 피자를 시켰으니까 좀 센스있는 곳이었다면 중복된다는 멘트 해주셨을 것 같은데 그 정도 서비스가 되는 곳은 아니었다. / 경복궁 근처 가면 한 번 가볼만한 곳 정도입니다. 재방문 의사는X". Therefore, it seems that we can roughly classify genuine and fake reviews by looking at the sentiment change of the reviews rather than the sentiment value of the whole review.

5. Conclusion

Fake reviews generated by GPT and genuine reviews collected from Naver Maps were pre-trained with three Korean language models to obtain sentence similarity, and classification was attempted using the K-medoids method. All three models did very well at catching fake reviews generated based on GPT, with recall values of 0.898, 0.871, and 0.931. When we further analyzed the clustering results of the KR-SBERT-V40K-klueNLI-augSTS model that best classified the test reviews, we found that the set of fake reviews generated by GPT tended to be heavily skewed toward the positive, with most of the sentiment changes in the reviews being positive-positive, positive-neutral, and neutral-positive. While it is still difficult to determine whether a review is real or fake based on a single review, it seems reasonable to assume that if you generate a set of reviews, and the set is a collection of reviews with a certain level of positivity and a wide range of sentiment, then it is likely to be a set of real reviews.

6. Future Research Plan & Proposal

In order to advance our research, we need to be able to classify the reviews in the grey zone, which seems very difficult with the current text-only approach. In order to take a text-based approach, we need data that is labelled as both real and fake, which is not currently available for Korean datasets. Translating the Yelp dataset into Korean and using it doesn't seem to be the answer either. This is because cultural differences make it seem impossible to resolve sentence structures like saying Korean Town or Chinese Town when talking about a Korean or Chinese restaurant, or saying the waiter's name and tipping for good service. While the quality of GPT-generated and translated reviews is good for generic reviews like "This food was delicious and I want to go again!" or for things that don't differ much from overseas, like steakhouses and Italian restaurants, it doesn't seem to be enough to cover restaurant reviews. As a possible direction, we should expand our test review data by getting more experimenters, and increase the proportion of labelled data as much as possible by checking whether the reviewer has written an advertorial review on NAVER Blog. In addition, we believe that new classification criteria should be found using additional information beyond the text, such as author information and the date the review was written.

7. Reference Literature

[1] Regina, R., Rini, E. S., & Sembiring, B. K. F. (2021). The Effect of Online Customer Review and Promotion through E-Trust on the Purchase Decision of Bukalapakin Medan City. International Journal of Research and Review, 8(8). https://www.ijrrjournal.com/IJRR_Vol.8_Issue.8_Aug2021/IJRR033.pdf

[2] 추천·보증 등에 관한 표시·광고 심사지침 개정. (2020). 공정거래위원회. Retrieved September 4, 2023, from https://www.ftc.go.kr/www/selectReportUserView.do?key=10&report_data_no=8597#:~:text=%EA%B3%B5%EC%A0%95%EA%B1%B0%EB%9E%98%EC%9C%84%EC%9B%90%ED%9A%8C(%EC%9C%84%EC%9B%90%EC%9E%A5%20%EC%A1%B0%EC%84%B1%EC%9A%B1,1%EC%9D%BC%EB%B6%80%ED%84%B0%20%EC%8B%9C%ED%96%89%ED%95%9C%EB%8B%A4.

[3] 사회관계망서비스(SNS) 부당광고 상시 점검 결과 발표. (2023). 공정거래위원회. Retrieved September 7, 2023, from https://www.ftc.go.kr/www/selectReportUserView.do?key=10&rpttype=1&report_data_no=9936

[4] 네이버 "방문자 리뷰" 조작 행위 성행, 도입 취지 무색. (2022, October 14). 뉴데일리. https://biz.newdaily.co.kr/site/data/html/2022/10/14/2022101400083.html

[5] 이민철&윤현식. (2020). 머신러닝을 활용한 가짜리뷰 탐지 연구: 사용자 행동 분석을 중심으로. 지식경영연구 21(3), 177-195.

[6] 김, 조, 이, & 정. (2021, January). 딥러닝을 이용한 배달 음식점 리뷰 자동 생성. 한국컴퓨터정보학회 동계학술대회 논문집, 29(1).

[7] Collective Opinion Spam Detection: Bridging Review Networks and Metadata. Shebuti Rayana, Leman Akoglu, ACM SIGKDD, Sydney, Australia, August 10-13, 2015

[8] Park, & Shin. (2021). KR-SBERT: A Pre-trained Korean-specific Sentence-BERT model. Github Repository. https://github.com/snunlp/KR-SBERT

[9] Gan. (2021). ko-sentence-transformers. Github Repository. https://github.com/jhgan00/ko-sentence-transformers

[10] Nils, & Iryna. (2019, November). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. http://arxiv.org/abs/1908.10084

[11] Ham, J., Choe, Y. J., Park, K., Choi, I., & Soh, H. (2020). Kornli and korsts: New benchmark datasets for korean natural language understanding. arXiv preprint arXiv:2004.03289

[12] Schubert, & Rousseeuw. (2021). Fast and Eager k-Medoids Clustering: O(k) Runtime Improvement of the PAM, CLARA, and CLARANS Algorithms. Information Systems, 101. https://doi.org/10.1016/j.is.2021.101804

[13] Lei Li, Yu-Ling Lan, Lei Huang, Congting Ye, Jorge Andrade, and Patrick C Wilson. "Selecting representative samples from complex biological datasets using k-medoids clustering". Frontiers in Genetics (2022) p. 1787. DOI: 10.3389/fgene.2022.954024

Supplementary Information

[Table S1] Number of Collected Reviews

Restaurant	Review Number	
	Raw Data	Processed Data
A	1662	837
B	1361	953
C	1648	1047
D	777	527
E	2996	1896
F	1928	1335
G	1220	1004
H	766	451
I	832	344
J	1347	840
Sum	14537	9234

[Table S2] Performance of K-Medoids Clustering

Model	No. Clusters	Loss	GPT-Review Recall	Test-Review Recall	Fake-Estimation of Real Reviews
KR-SBERT-V40K-klueNLI-augSTS	34	1795	0.898	0.8	0.410
ko-sroberta-multitask	35	1110	0.871	0.75	0.365
paraphrase-multilingual-MiniLM-L12-v2	36	1062	0.931	0.7	0.366

[Table S3] Test Review Classifying Result

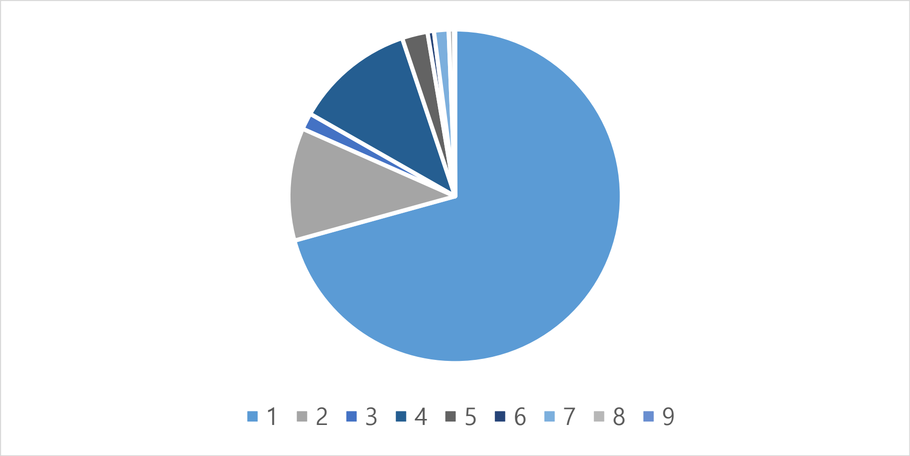
Review Text	KR-SBERT-V40K-klueNLI-augSTS	ko-sroberta-multitask	paraphrase-multilingual-MiniLM-L12-v2
1. 새로 생긴 식당이라 그런 지 인테리어랑 식기도 깔끔하고, 음식도 깔끔하게 플레이팅 되어 나왔어요	Real	Real	Fake
2. 첫 방문이었는데 사람이 많아서 그런 지 음식 나오는데 20분 정도 걸렸다 ㅠㅠ 그럼에도 너무 자극적이지도 않으면서도 음식 간이 적당해서 좋았다	Real	Real	Real
3. 명란오일파스타 시켰는데 명란 맛이 잘 안 느껴졌어요 ㅠㅠㅠ	Real	Real	Real
4. 카츠나베 시켰는데 위에 올라간 반숙란이랑 고슬고슬한 쌀밥에서 비주얼부터 먹음직... 돈카츠랑 새우카츠 반숙란과 국물이랑 같이 먹으니 너무 잘 어우러졌다	Real	Real	Real
5. 서울에서 줄 서서 먹는다던 호호식당이 대전에도 생겼다길래 가봤다 ㅋㅋ 연어 메뉴가 다 품절이라 아쉬웠지만 아쉬운대로 파스타랑 돈카츠 메뉴 시켰는데 둘 다 기대 이상으로 맛있었음	Real	Real	Real
6. 친구가 유명한 곳이라고 추천해서 가봤는데 유명한 이유가 있더라구요!!! 파스타는 제 취향이 아니었는데 밥이 진짜 대박이었어요. 찰기 가득하고 윤기가 흐르는게! 재방문 의사 완전 있음!!	Fake	Fake	Real
7. 주말 저녁에 방문해서 20분 가량 줄서서 기다렸는데 음식이 깔끔하고 정갈하게 나와서 만족했습니다.	Real	Real	Fake
8. 대전에서 연어 스테이크 파는 곳이 몇 없는데 그걸 여기에서 팔더라고요. 그거 하나만으로도 다시 올 생각입니다.	Real	Real	Real
9. 가츠나베가 정말 맛있었어요. 간장 베이스 양념이 잘 스며들어서 너무 부드러웠습니다.	Real	Real	Fake
10. 역시 백화점에 있어서 인테리어랑 플레이팅이 깔끔하게 나오더라고요. 음식 맛도 정갈하고요. 백화점 왔다가 마무리로 깔끔한 식당입니다.	Real	Real	Fake
11. 탕수육이 정말 맛있음. 고기가 매우 부드럽고 튀김 옷 두께도 적당해서 부드러운 식감으로 먹기에 매우 좋음. 유니짜장은 소스 간은 단맛이 적고 슴슴한 편이라 호불호가 강할 수 있을 듯함. 다만 면이 적당히 가늘고 부드러워서 소스와의 조화를 생각했	Fake	Real	Fake

을 때는 딱 알맞는 것 같음.			
12. 동파육 존맛탱.... 전반적으로 맛있어요~~~	Fake	Real	Real
13. 가격대가 좀 있지만 양 많고 맛있어요 ㅠㅠㅠㅠ	Real	Fake	Real
14. 탕수육이랑 유니짜장면 시켰는데 탕수육이 진짜 포슬포슬하고 맛있어요!! 강추합니다!!	Fake	Real	Real
15. XO볶음밥 대전 다른 곳에서 잘 안 파는데 여기서 팔아서 너무 좋았어요! 다만, 짬뽕은 대전 다른 집에 비해서는 그냥 무난무난한 맛...	Real	Real	Fake
16. 맛있는 중국집이라고 해서 듣고 갔는데 나쁘진... 않았는데 가격은 조금 비싼 느낌이 있네요 ㅠ	Real	Fake	Real
17. 탕수육은 맛있었어요. 그런데 짜장면은 그닥이었어요. 대전에는 맛있는 짜장면집 찾기가 힘드네요.	Real	Fake	Real
18. 음식은 맛있었는데 주말 점심이라 그런지 가게도 너무 혼잡한 느낌이 있었습니다.	Real	Real	Real
19. 친구들이랑 여럿이서 갈때 괜찮은 것 같아요. 가게가 커서 6명이서 앉을 수 있는 테이블도 많았고, 탕수육 하나 시켜서 나눠먹기 딱 좋네요.	Real	Real	Fake
20. 짬뽕이 맛은 있는데 다른 갈마짬뽕이나 이비가짬뽕 같은 곳이랑 비교하면 가격이 비싸서 좀 아쉽네요.	Real	Fake	Real

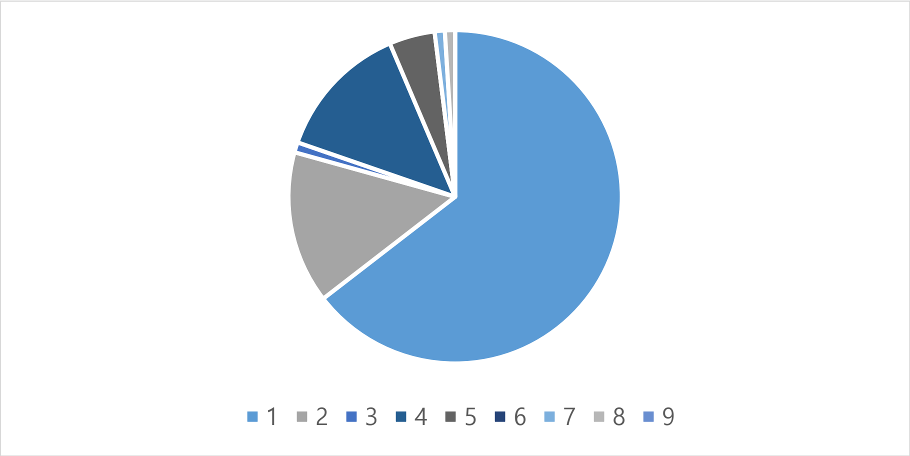
[Table S4] Average Sentiment of Clusters

클러스터 #	Negative	Positive	Neutral
3	1.931518	95.84795	2.220536
24	3.184265	93.13084	3.684896
27	54.45527	37.40738	8.137351
30	19.12387	79.23171	1.64442
10	6.000033279	90.13564333	3.864322395

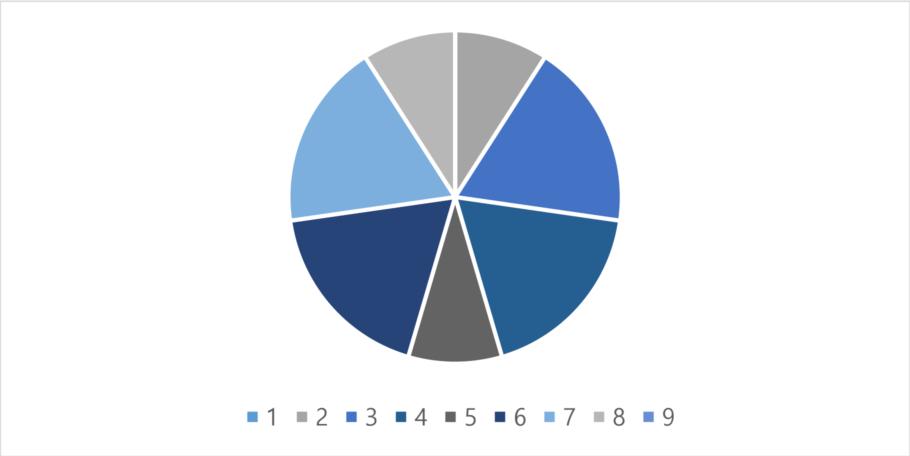
[Table S5] Sentiment Flow of Cluster#3



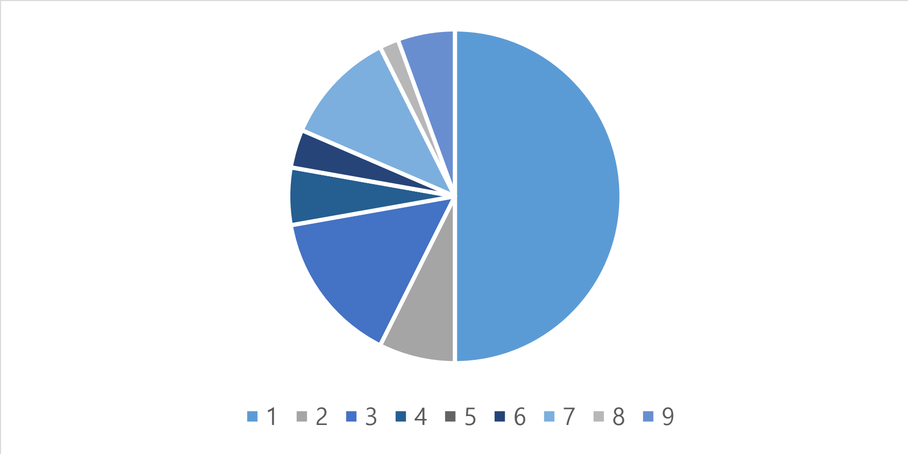
[Table S6] Sentiment Flow of Cluster#24



[Table S7] Sentiment Flow of Cluster#27



[Table S8] Sentiment Flow of Cluster#30



[Table S9] Sentiment Flow of Cluster#10