



Using **Naver** review data

Restaurant Recommendation System

Based on Review Data

202001712 Seo Minseok

202000318 Kwon Soonjae

202102670 Lee Jeongseok

202002439 Lee Sangmin

Table of Contents

1. Introduction to the Topic

2. Data Collection and Preprocessing

3. Creating a Base Model Using Four Methods

- Transformer
- Keyword-Based Filtering(TF-IDF)
- Keyword-Based Filtering(Word2vec)
- BERT

4. Adjusting Project Direction After Faculty Feedback

- Topic Title Revision
- Simplified Version of the Review Dataset

5. Review Text Summarization Model

6. Review Data Preprocessing, Version 2

7. Model Development

- Model 1
- Model 2

8. Project Evaluation

- TF-IDF Score
- BERT Score

Introduction to the Topic



What Kind of Restaurant Should We Choose?

Data_Reflecting_Naver_Reviews
Specific_Requirements
Companion_Information
Cuisine_Category
Location
Waiting_Time_Information

Difficulties in Decision-Making Based on Ratings
Time-Consuming to Find Relevant Reviews



Restaurant Recommendations Tailored to User Preferences



Contemplating....

Data Collection and Preprocessing



Data Collection and Preprocessing

데이터 수집과 전처리

<Columns After Preprocessing>

"Waiting Time," "Companion," "Review," "Store Name,"
"Address," "Category"

1



Data Collection Using the **Listly Program**

A Chrome extension program used to extract desired information directly from web pages. Exported as Excel files for further processing.

2

Data Preprocessing

Retain only **the necessary information** and remove irrelevant data. Consolidate the **remaining data** into the **same columns**.

Data Collection and Preprocessing

데이터 수집과 전처리

After organizing the columns

대기시간	동반자	리뷰	가게명	주소	가게분류
대기 시간 바로 입장	부모님	7시예약 했는데 일정이 빨리 끝나서 6시로 변경 요청 드렸는데 바로해주셨어요~ 맛있고 한적 쾌적 친절 했어요 엄마랑 조카가 엄지척♡ 맛있다며 계속	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친구	파스타 밖에서 잘 안사먹는 편인데 너무우우 맛있게 먹었어요! 메뉴 하나하나 다 특색있게 다르게 맛있어요ㅠㅠ 특히 청양 엔조비 오일 파스타는 두고두고 생각날 거 달아서 또 먹으려 오려구요,, ❤️ 다 맛있어서 도장깨기 하고 싶습니다ㅜ 도전! 뷔도 너무 좋아서 날씨 좋은데 청량한 기분 느끼면서 맛있게 먹었습니다😋 데이트하기에 최고네용!! 아 그리고... 사장님 예뻐요...😊 다른 메뉴도 맛보러 (사장님 얼굴 보러) 또 갈게요~~😊😊	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친구	음식이 진짜 대박이예요!!! 시킨 메뉴 다 맛있고 플레이팅도 너무 예뻐서 사진 찍을 맛 나요! 직원 분들도 다 친절하시고 사장님 예쁘고 좋아요❤️ 성수역 데이트, 나드리, 청모 등 어떤 모임이든 다 강추요! 🤗🤗😊😊	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	아이	트러플 향이 너무 좋아요~ 음식이 대체적으로 맛있고 너무 친절하셔서 즐거운 식사했습니다.	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친구	분위기가 좋고 음식도 맛있어요! 브런치치도 너무 맛있고 대박입니다:-) 또 올래여	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	지인 · 동료	음식 정말 맛있네요 곧 재방문 할꺼예요! 먹어본 메뉴 도장깨기 할려구용^^!!	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친척 · 형제	두 가지 다 너무 맛있어요! 브런치는 잼, 휘핑, 시럽 곁들여 먹을 소스가 많아서 좋았고 안에 있는 빵이나 소시지 등은 따로 추가해먹을 수 있는게 좋은 것 같아요! 명랑파스타 맛집 찾고 있었는데 여기에요! 해산물도 크고 도톰하게 많이 들어가 있고 느끼하지 않고 고소한맛!	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	부모님	부모님 모시고 간단히 먹을 곳 찾다가 들어왔는데 테이블끼리 간격도 넓고 음식도 빨리 나와서 딱 좋았어요. 친구들끼리 와도 좋을 것 같은데, 여기 라구 파스타가 진짜 요물이네요. 매콤해서 살짝 느끼할 수 있는 브런치 메뉴랑 잘 어울려요. :)	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	지인 · 동료	아메라칸스타일 브런치가 맹거져 찾아갔어요~ 찾고았던 딱 그 맛이였어요! 시간이 많이 없어서 빨리 내달라고 부탁드렸더니 상황에 맞게 넘 친절하게 대응해주시고 명함이벤트도 하길래 명함도 넣고 와쏘요~ 레몬에이드도 달지않고 건강한 맛이였어요 또 갈게요	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	지인 · 동료	으앙너무 맛나고 몽블랑파스타 미쳤어요 신메뉴라고하는데 너무 꼬숩고 든든하게 먹고갑니다 오랜만에 친구만났는데 친구도 너무 좋다고 했습니다 자리도 너무쾌적해서 나중에 데이트나 소개팅 장소로 엄청 추천드립니다😊😊😊 루꼴라쳐들이인데 루꼴라도 아낌없이 넣어주셔서 감사합니다	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친구	오랜만에 모임으로 시그니처 성수 방문했어요~ 직원분도 친절하시고 식사도 맛있습니다. 예약하고 가는게 좋을 것 같아요!	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	친구	친구들이랑 생일 기념했는데 분위기 좋아요! 스테이크도 가성비 좋은것같고 얼그레이토스트??도 맛있었어요:) 굿굿	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식
대기 시간 바로 입장	부모님	양이 많고 음식이 빨리 나와서 너무 좋았습니다! 맛도 맛있구 식당 분위기도 차분하고 좋았어요!	37.5 시그니처 성수	서울 성동구 연무장길 41-19 1,2,3층 37.5 시그니처 성수	양식

Creating a Base Model Using Four Methods



Base Model

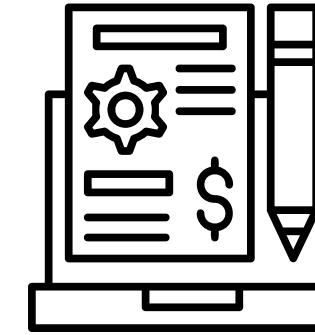
*Using Four Models
Find and Recommend Reviews Similar to the Input*

Model	Common Preprocessing	Content Analysis	Output
Keyword-Based Filtering TF-IDF		Faced issues as increased reviews led to smaller IDF values, resulting in low similarity scores → Applied BM25 algorithm for better similarity checks and showed matching reviews.	
Keyword-Based Filtering Word2vec	Tokenization, Morphological Analysis, Stop word Removal	Sentiment analysis using positive/negative keyword lists ; calculate scores for each aspect (e.g., waiting time, companions) and compare scores to find the top-ranked reviews.	Similarity score, sentiment analysis score, estimated waiting time, category parsing.
BERT		Analyzed data (store name + address + category + average waiting time + review score + numerical values); output included combined scores based on inputs.	

Adjusting Project Direction After Faculty Feedback

Adjusting Project Direction After Faculty Feedback 1

교수님과의 면담이후 프로젝트 방향성 수정



1. Topic Revision

Original Topic

" Chatbot-Based Restaurant
Recommendation System Using
Review Data "



Revised Topic

" Restaurant Recommendation System
Based on Review Data "

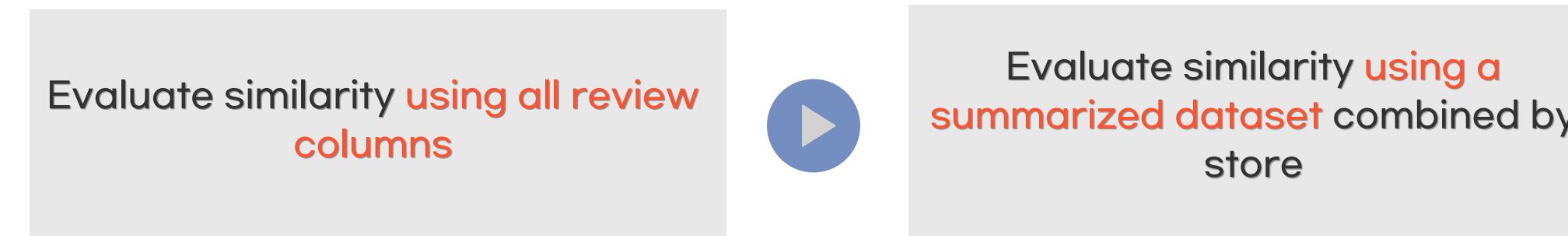


A system that provides **one-time restaurant recommendations** based on simple input is **limited in being labeled as a "chatbot."** Therefore, it was revised to a "**Restaurant Recommendation System.**"

Adjusting Project Direction After Faculty Feedback 1

교수님과의 면담이후 프로젝트 방향성 수정

2. *Converting the Review Dataset into a Summarized Version*



바꾼 원인

Cause

- Using the existing dataset may result in restaurant recommendations overly biased toward specific reviews.
- Therefore, it was concluded that a comprehensive review dataset for each restaurant should be created for more balanced evaluations.

결과

Effect

- Grouped reviews by store and combined them to form a new dataset with summarized texts.
- Since the total length of reviews (approximately 200 per store) was too long, the texts were divided into manageable sections based on a token count (1024 tokens) and summarized iteratively.
- As a result, a single summarized review was created for each restaurant, enabling comprehensive evaluations.

Review Text Summarization Model



Review Data Summarization Model Implementation

● Objective

Summarize reviews for each store to enable efficient analysis and utilization using the KoBART model.

● Key Steps

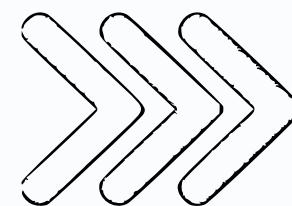
- Data Preprocessing: Remove special characters, deduplicate, and refine repetitive patterns.
- Review Grouping: Group reviews to ensure token length does not exceed 512.
- Summary Generation: Use beam search to generate summaries for each group (max 200 tokens).
- Recursive Summarization: Integrate group summaries into a final consolidated summary.

● Code Components

- KoBART model initialization using [gogamza/kobart-base-v2](#).
- Implementation of text cleaning and deduplication functions.
- Summarization of grouped review data with KoBART.
- Exporting results to an Excel file.

Review Data Summarization Model Implementation

지나가다가 우연히 들어왔는데 양념갈비들이 고기가 너무 퀄리티가 좋아요. 반찬도 다양하게 나오고 맘에들어요. 초등학생 아이가 여기 고기 너무 부드럽대요. 흰라면 특이하게 고수 들어가서 좋네요(개인적으로 고수러버) 또 올게요 흑백요리사 출연식당이라고해서 가봤어요 ㅎㅎ 요리는 방송에 한컷도 안나오긴 했지만 사장님 얼굴이 딱 한장면 나오신거 귀엽 ㅎㅎ ㅎ고기는 가성비 좋고 평범한 맛이에요! 양은 많아요! 적당히 가기 괜찮을 듯합니다! 그리고 버터 나와서 발라먹으니 맛있네요! 서비스가 진짜 계속 나와요!!!! 해장국 냉면 식혜 다 넘 맛있게 먹었구요, 직원분과 점장님께서 진짜 친절하셔서 기분 좋은 식사였습니다 고기도 양 너무 많고 맛있어서 오길 잘했다고 계속 그랬습니다 다음에 또 올게요 처음 왔는데 맛있습니 다. 재방문 의사 있습니다! 요즘 핫한 남영동 양문!!! 드디어 저도 가봤는데 요역시나 핫하고 맛있어요 ㅎㅎ 추천!!! 최근에 먹은 고깃집 중에 가장 맛있는 곳이었어요... 소갈비살이 전혀 질기지도 않았고, 잡내도 안났구요! 기본 밑반찬도 다양하게 나오고 고기 엄청 들어간 해장국?까지 서비스라고 하셔서 솔직히 놀랬네요.. 진짜 맛있게 잘 먹었고, 다음엔 가족들과 같이 오려구요!! 아 진짜 맛있었어요 모든 메뉴가 다-!! 저의 또또또 또또간집이에요! 리뉴얼 전부터 근처에서 고기 먹을 때 여기만 왔는데 리뉴얼 되고 나서 마늘 대파 등 양념이 더 맛있어져서 이틀 연속 왔어요! 강추 가성비 끝판왕 맛집입니다. 강력 추천 생갈비 신선도 퀄리티 무슨일.. 때깔 미친만큼 맛도 너무 좋습니다. 기본찬도 너무나 훌륭해서 서울에 살았다면 단골하고픈 맛집이에요 b 소갈비 저렴한 가격에 맛있게 먹었어요. 사이드 메뉴도 양이 많아서 엄청 배부르네요. 재방문 예정입니다! 소갈비 맛있.....



진짜 너무 맛있어요 한남동에 있는 양문보다 맛있네요 ㅎ 밑반찬도 맛있고 고기랑 막창 다 존맛이었습니다 최고 서비스가 해장국 비빔면 양념게장 등 미쳤어요 지나가다가 우연히 들어왔는데 양념갈비들이 고기가 퀄리티가 좋아요 반찬도 다양하게 나오고 맘에들어요. 초등학생 아이가 여기 고기 부드럽대요 흰라면 특이하게 고수 들어가서 좋네요 개인적으로 고수러버 또 올게요 압구정 찐 맛 집 발견했어요 소고기 정말 신선하고 불에 구워서 같이 주시는 버터 위에 발라 먹으면 고기에 남은 열때문에 버터가 녹아 스며들어요 입에서 없어 없어지는 맛입니다 시킨 라구파스타는 불판위에 구우면서 계란물이랑 익혀먹는 게 신기했어요 안에 고소한 신기 좋아갈비 한판에 육 잘 소.

Review Data Preprocessing, Version 2

Dataset Modification

Column Name

"동반자"

Additional Column Information

"동반자 1순위", "동반자 2순위"
"동반자 1 순위 비율", "동반자 2순위 비율"

Column Name

"대기시간"

Additional Column Information

"대기시간 1순위", "대기시간 2순위"
"대기시간 1 순위 비율", "대기시간 2순위 비율"

Column Name

"리뷰"

Additional Column Information

"요약 리뷰"

Model Development

Implementation of Two Recommendation Models

두가지 추천 모델 구현

Two Models for the Recommendation System

Model 1

"Similarity-Based Priority"

- Using **SBERT** and **cosine similarity** to recommend the most relevant restaurants based on the user's query.

Model 2

"Column-Based Information Priority"

- Although SBERT similarity was implemented in the same way, the system analyzes user **input text to extract information** such as waiting time, restaurant category, companion details, and restaurant address, and uses similarity along with other data to **evaluate each factor in a balanced manner**.

Model 1

Purpose of Column Usage

컬럼 사용 목적

Column Name	Details	Purpose
가게명	Refers to the name of the restaurant mentioned in the review.	Provides the restaurant name for recommendations.
주소, 요약주소	<ul style="list-style-type: none"> - Address: Refers to the actual location name. - Summary Address: Refers to a specific area or region. 	<ul style="list-style-type: none"> - Summary Address: Provides the location where the user wants to find restaurants, allowing exploration of restaurants in that area. - Address: Provides location information for recommendations.
동반자 1순위, 동반자 2순위 동반자 1 순위 비율, 동반자 2순위 비율	Categories for companions (e.g., date, spouse, alone, family, children, friends); checks their frequency and calculates the ratio for Top 1 and Top 2 companions.	If the companion category requested by the user matches the first or second priority companion category, multiply the companion ratio by 0.1 to calculate the weight.
대기시간 1순위 대기시간 2순위 대기시간 1순위 비율 대기시간 2순위 비율	Categories for waiting times (e.g., immediate entry, 10 minutes, 30 minutes); checks their frequency and calculates the ratio for Top 1 and Top 2 waiting times.	Purpose of Providing Information for Recommendations
가게 분류	Provides information on restaurant types (e.g., Korean, Western, Japanese, Chinese, Asian fusion).	Supplies category information to recommend restaurants that match the user's preferred cuisine.
리뷰	Summarizes all reviews for a restaurant and provides an overall evaluation.	Uses SBERT models to embed review data and user input into vectors, calculating cosine similarity between the input and summarized review embeddings.

Model 1

애인이랑 갈건데 성수에 있는 스테이크가 맛있는
양식집 추천해줘

User

Output

평균 리뷰 유사도: 0.6276

평균 최종 유사도: 0.6276

추천된 가게별 유사도:

가게명: 문화식당(양식)

주소: 서울 성북구 혜화로 88 2층, 3층

리뷰 유사도: 0.6675

최종 유사도: 0.6675

48.9%의 사람들이 친구와 함께 왔습니다.

많은 사람들이 0.0분의 대기시간을 가졌습니다.

가게명: 맥시크(양식)

주소: 서울 성동구 상원1길 35 1층

리뷰 유사도: 0.6279

최종 유사도: 0.6279

39.4%의 사람들이 친구와 함께 왔습니다.

많은 사람들이 0.0분의 대기시간을 가졌습니다.

가게명: 놀끼룸(양식)

주소: 서울 성동구 성수일로1길 8 1층

리뷰 유사도: 0.5874

최종 유사도: 0.5874

42.5%의 사람들이 연인 · 배우자와 함께 왔습니다.

많은 사람들이 0.0분의 대기시간을 가졌습니다.

1 Keyword Extraction and Filtering

Extracts keywords such as companions, summary addresses, and restaurant categories from user input to **filter restaurants**.

2 Embedding and Similarity Calculation

Embeds summarized reviews and user input into **vectors using the SBERT model**, then **calculates cosine similarity** between user input and the review data embedding vectors.

3 Restaurant Recommendation

Recommends restaurants based on companion weights and cosine similarity.

Model 2

Purpose of Column Usage

컬럼 사용 목적

Column Name	Details	Purpose
가게명	Refers to the name of the restaurant mentioned in the review.	Provides the restaurant name for recommendations.
주소, 요약주소	- Address: Refers to the actual location name. - Summary Address: Refers to a specific area or region.	- Summary Address: Provides the location where the user wants to find restaurants, allowing exploration of restaurants in that area. - Address: Provides location information for recommendations.
동반자 1순위, 동반자 2순위 동반자 1 순위 비율, 동반자 2순위 비율	Categories for companions (e.g., date, spouse, alone, family, children, friends); checks their frequency and calculates the ratio for Top 1 and Top 2 companions.	Provides information on companion types preferred by the user, and assigns weighted scores based on the user's selected companion ratios.
대기시간 1순위 대기시간 2순위 대기시간 1순위 비율 대기시간 2순위 비율	Categories for waiting times (e.g., immediate entry, 10 minutes, 30 minutes); checks their frequency and calculates the ratio for Top 1 and Top 2 waiting times.	Provides information about the types of waiting times preferred by the user and assigns scores by applying weights to the ranking ratios of the selected waiting times. Provides waiting time information during recommendations.
가게 분류	Provides information on restaurant types (e.g., Korean, Western, Japanese, Chinese, Asian fusion).	Supplies category information to recommend restaurants that match the user's preferred cuisine.
리뷰	Summarizes all reviews for a restaurant and provides an overall evaluation.	Stores the text entered by the user multiple times, calculates SBERT similarity scores, multiplies them by 0.5 , and assigns points based on the results.

Model 2

모델 2 설명

User

가족과 식사를 하고싶어. 좀 깔끔하고 분위기 좋은 식당 추천해줘.

대기시간은 어느정도가 괜찮으실까요? (예: 10분, 30분, 즉시 입장, 없음, 대기시간 없음)

User

즉시 입장이 좋겠지?

어떤 종류의 음식을 먹고싶으세요? (예: 일식, 한식, 중식, 양식, 아시안)

User

양식을 먹고싶어. 이쁜하게 익은 스테이크와, 간이 타이트하고 면의 익힘정도도 적당한 파스타를 잘하는 집을 추천해줘봐.

어느 지역에서 찾고 계신가요? (예: 성수, 건대, 왕십리, 압구정)

User

성수에서 먹을거야~

네. 요구사항에 맞추어 음식점 세곳을 추천해드릴게요.

Model 2

모델 2 설명

Stores the category and input text.

Input_data				
✓	0.0s			
동반자	대기시간	가게분류	주소	
0	가족	대기 시간 바로 입장	양식	성수

< Stores the Entire Sentence Entered by the User >

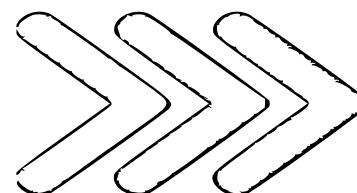
"

가족과 식사를 하고싶어. 좀 깔끔하고 분위기 좋은 식당 추천해줘. 즉시 입장이 좋겠지? 양식을 먹고싶어. 이쁜하게 익은 스테이크와, 간이 타이트하고 면의 익힘정도도 적당한 파스타를 잘하는 집을 추천해줘봐. 성수에서 먹을거야~

"

- Address, Restaurant Category:
Only selects matching restaurants.

- Waiting Time, Companion:
Calculates weighted scores and converts them into points.



- User Input:
Evaluates similarity with review data and converts it into a score.

Recommendation Results

추천 음식점 리스트:

1. 엘로코 성수 (양식)

주소: 서울 성동구 동일로 75-18 1F, 2F
이 가게의 방문자는 51%가 '연인 · 배우자'였습니다.
대기시간은 97%가 '대기 시간 바로 입장'로 나타났습니다.
유사도 점수: 0.77
최종 점수: 0.930

2. 리틀포레스트 (양식)

주소: 서울 성동구 성수일로12길 23 2층
이 가게의 방문자는 57%가 '친구'였습니다.
대기시간은 100%가 '대기 시간 바로 입장'로 나타났습니다.
유사도 점수: 0.70
최종 점수: 0.922

3. 리타르단도 (양식)

주소: 서울 성동구 뚝섬로 393 B1 리타르단도
이 가게의 방문자는 51%가 '친구'였습니다.
대기시간은 99%가 '대기 시간 바로 입장'로 나타났습니다.
유사도 점수: 0.74
최종 점수: 0.920

Project Evaluation

TF-IDF Score

	리뷰_원본 \ 가게명
0	37.5 시그니처 성수 7시예약 했는데 일정이 빨리 끝나서 6시로 변경 요청 드렸는데 바로해주셨어요~ 맛있고...
1	GTS 포레스트 버거가 되게 양많고 맛잇어요! 감자튀김도 엄청 바삭하고 맛잇습니다. 강추합니다. 분위...
2	Just Steak 스테이크 맛집~ 예약 했고 와인과 샴페인은 가져갔어요. 파스타 구운야채 모두 맛있고 티...
3	Lu307 이자카야 불빛이 좀 노란빛이라 음식색감이 잘 안나와 아쉽지만 승승그래도 회 매우 신선하고 단...
4	Mr박의돈워리 친절해서 이용하기 좋습니다 이제부터 돼지고기는 여기야장이 허락되는 요기간놓치지마시고...
..	...
176	홍능족발 튀김족발 처음 먹는데 진짜 맛있어요 😊 예전에 오고 맛있어서 고양시에서 또왔어용 튀김...
177	홍성집 정갈하고 정성있는 한식집으로 가족과의 모임에서 강력 추천해요 ❤️ 한우 육회...
178	화기애애 본점 맛있어요~~~~~ 옛날부터 왔던곳입니다 🎂 몇년을 와도 변함이 없네요 🍷 건대...
179	회식의달인 건대점 고기질이 넘 좋아요!! 소고기는 프라임등급이라 그런지 냄새 하나도 안나고 야들야들 부...
180	희릿 음식이 맛있고 분위기가 좋아요 친절합니다 분위기 너무 좋아서 예전부터 오고 싶었는데 ...
	리뷰_요약 TF-IDF 유사도
0	친구 오셔서 행복한 시간 보내세요 개인적으로 친구랑 즐거운 저녁식사 다들 리조또... 0.258526
1	gts버거 너무 맛있어요 육즙가득 사이드도 다양하구 양도 딱 좋아여 허 재방문각 가게... 0.274196
2	어서 많이 없어서 딱 스테이크 전문점이란 느낌이 들었습니다 스테이크 너무 주변사람들에... 0.184906
3	너무 시끄러워요 분위기도 좋고 맛있고 친절하셨어요 일차만 먹었는데도 배도 엄청 불... 0.186241
4	하게 보내면서 맛있는 고기는 덤~ 뷔도 좋아요~ 너무 친절해서 더욱 좋아요~ 여기 고기... 0.302675
..	...
176	쫄깃쫄깃하고 맛있어요 사이드 메뉴 된장찌개 맛있어요. 쟁반국수 족발에 싸먹으니 아주 ... 0.277459
177	음식도 너무 맛있고 직원들도 친절하시고 짠당골 할했습니당 맛있습니다 재방문 할듯요 맛있게 잘먹... 0.136302
178	많이 많이렸습니다 토요일이라 오래 기다렸는데 오늘도 넘 맛잇었다 고기 굽는데 좀 오래... 0.208603
179	가족 많이 잘먹었습니다 감사 합니다 번창하세요 소고기도 잡내없이 맛있고 특히 등심이... 0.177434
180	도 많이 적당하고 좋습니다 지나가다 들어왔는데 분위기가 좋아요 야외테이블 예약하고 ... 0.232838
..	[181 rows x 4 columns]
	max: 0.3957819629519678
	min: 0.023385135494472426
	mean: 0.18947676033912464

- Since the summarization is abstractive, the TF-IDF score tends to be relatively low.
- TF-IDF is a method that quantifies textual data to calculate the importance of words.
- This method heavily depends on the occurrence of individual words, thus failing to reflect semantic similarity or word rearrangement.

BERT Score

```
BERTScore Precision:  
max : 0.8218975067138672  
mean : 0.715890645980835  
min : 0.6317193508148193  
  
BERTScore Recall:  
max : 0.7463638782501221  
mean : 0.6586269736289978  
min : 0.5475931167602539  
  
BERTScore F1:  
max : 0.7823116779327393  
mean : 0.6858066320419312  
min : 0.5919321179389954
```

● Precision

Precision represents the proportion of the summary that semantically matches the original text.

An average score of 0.7159 indicates that **the summary aligns well with the original text**, and the minimum score (0.6317) is not significantly low, showing that **most summaries include the meaning of the original text**.

● Recall

Recall measures how much of the meaning in the original text is preserved in the summary.

An average score of 0.6586 shows that most information in the original text is retained in the summaries, but the minimum score (0.5476) suggests **that some summaries may have omitted certain information**.

● F1

F1 represents the balance between Precision and Recall. An average F1 score of 0.6858 indicates that the summaries are **well-matched to the original text in terms of semantics**. However, **the minimum score (0.5919) suggests that some summaries may not align well with the original text**.

Reasons for Not Using BLEU and ROUGE as Summarization Metrics

BLEU calculates precision, and ROUGE calculates recall. While they differ in focus, both methods rely on the alignment of n-grams, which may not provide accurate evaluations for abstractive summarization.

Additionally, the significant length difference between the summary and the original text may distort the results.

For these reasons, these two evaluation metrics were not used.

THANK YOU

감사합니다