

PROJECT 1.

ESG Classification by Implicit Rule Learning via GPT-4

01

ABOUT PROJECT

- ESG 관련 기사의 IMPACT TYPE과 IMPACT DURATION 예측
- LLM 모델을 활용한 ESG 요소 정보 분석

ESG Classification by Implicit Rule Learning via GPT-4

Title: ESG Classification by Implicit Rule Learning via GPT-4

Authors: Yun Hyojeong, 찬영 김, Moonjeong Hahm, Kyuri Kim and Guijin Son

ESG Classification by Implicit Rule Learning via GPT-4

Anonymous submission

Abstract

Environmental, social, and governance (ESG) factors are widely adopted as higher investment return indicators. Accordingly, ongoing efforts are being made to automate ESG evaluation with language models to extract signals from massive web text easily. However, recent approaches suffer from a lack of training data, as rating agencies keep their evaluation metrics confidential. This paper investigates whether state-of-the-art language models like GPT-4 can be guided to align with unknown ESG evaluation criteria through strategies such as prompting, chain-of-thought reasoning, and dynamic in-context learning. We demonstrate the efficacy of these approaches by ranking 2nd in the Shared-Task ML-ESG-3 *Impact Type* track for Korean without updating the model on the provided training data. We also explore how adjusting prompts impacts the ability of language models to address financial tasks leveraging smaller models with openly available weights. We observe longer general pre-training to correlate with enhanced performance in financial downstream tasks. Our findings showcase the potential of language models to navigate complex, subjective evaluation guidelines despite lacking explicit training examples, revealing opportunities for training-free solutions for financial downstream tasks.

Keywords: Large Language Model, Benchmark, Finance

1. Introduction

In recent years, there has been a noticeable increase in investors factoring environmental, social, and governance (ESG) considerations into

to generate realistic datasets. Moreover, the creation of large-scale, high-quality labeled datasets is resource-intensive. Manually annotating extensive text collections requires considerable time and skilled professionals. Furthermore, the accurate

작업 기간

2024.01 - 2024.05(예정)

팀원 구성

5명

프로젝트 TASK

ESG 뉴스 기사의 IMPACT TYPE과 IMPACT DURATION을 정확하게 분류

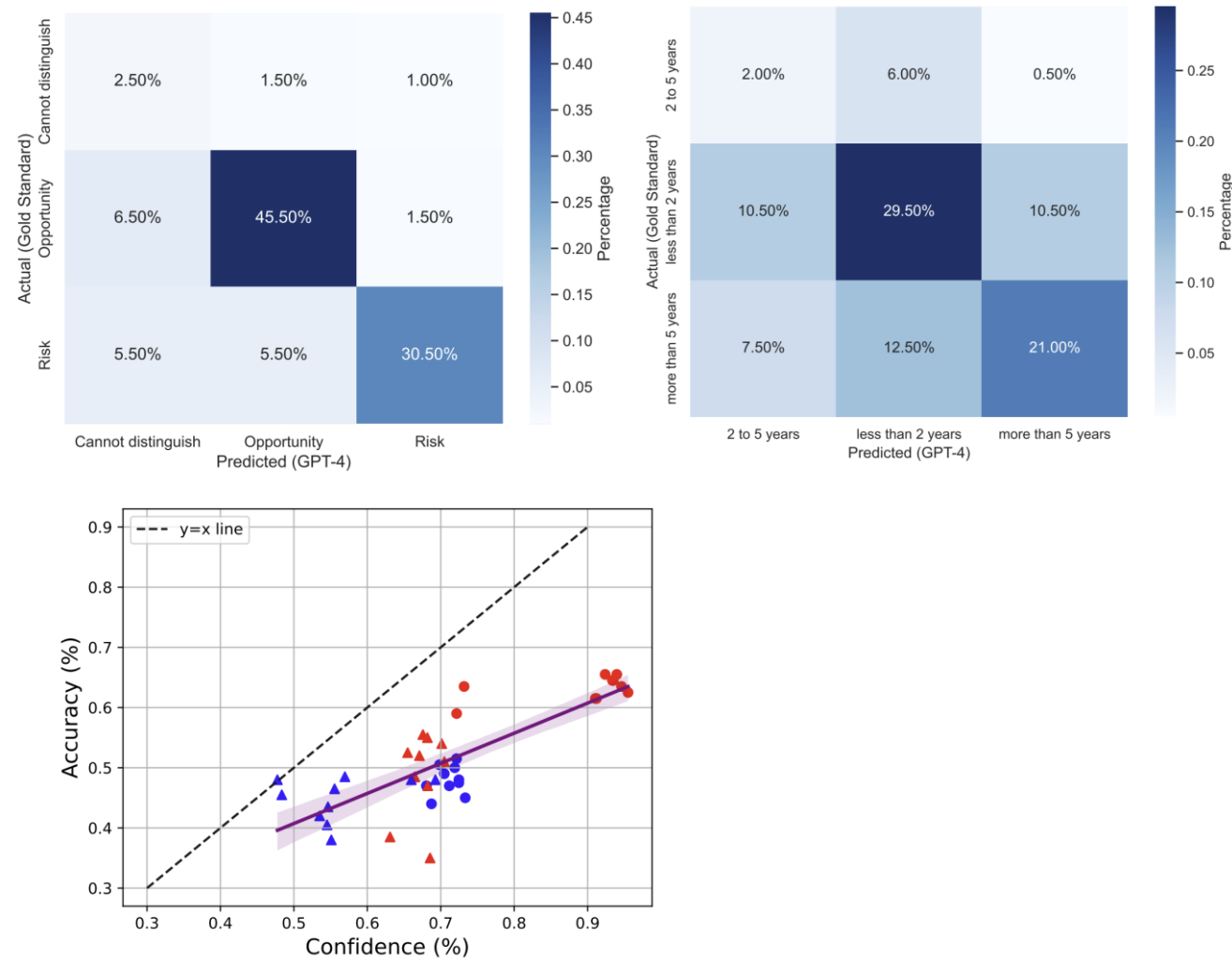
프로젝트 필요성

ESG 정보가 투자에 많이 사용되고 있으나, 평가 기관의 방법론은 지적 재산이기 때문에 공개되지 않는다. 더불어 초기 언어 모델은 구하기 어려운 학습 데이터가 필수적이며 분야 특성상 데이터 회전 속도도 높다.

프로젝트 내용

대규모 언어모델 (GPT-4)로 zero-shot question answering을 통해 비용문제 해결
GPT-4를 TASK에 더 적합하게 적용하기 위해 prompting, chain-of-thought, dynamic in-context learning (RAG) 사용

ESG Classification by Implicit Rule Learning via GPT-4



프로젝트 결과

한국어 데이터 세트 IMPACT TYPE 예측 2위

3월 19일 FINNLP-2024 에서 논문 ACCEPT

3월 27일 최종 수정 후,
5월 21일 이탈리아에서 워크샵 예정

PROJECT 2.

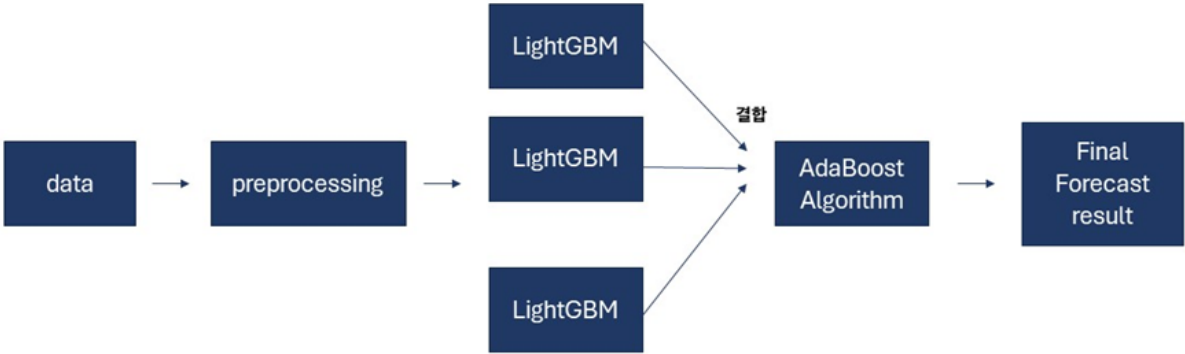
LSTM, AdaBoost-LGBM, CatBoost 활용, 초개인화 사고율 실시간 예측

02

ABOUT PROJECT

- 합리적인 보험료 조정을 통한 자동차 보험 시장에서의 점유율 상승
- 초개인화를 통한 고객별 사고율 예측을 기반한 보험 상품의 필요성
- 환경 변화에 따른 보험사의 실시간 대응 가능

LSTM, AdaBoost-LGBM, CatBoost 활용, 초개인화 사고율 실시간 예측



▲ 사고율 예측 모델 플로우

작업 기간 2023.08 - 2023.10 (3달)

팀원 구성 4명

프로젝트 목적

자동차 보험 고객 데이터를 활용한 고객별 사고 발생률 예측
및 보험산업 대응 아이디어 제안

프로젝트 내용

환경 변화 시 조기 전략 재수립이 가능한 가정 가변성을 확보하고 고객별 초개인화된 위험도 측정이 가능한 사고 예측 모델을 구축. 자연어 처리를 활용한 보험산업 대응 아이디어 제안

**주요 업무 및
상세 역할**

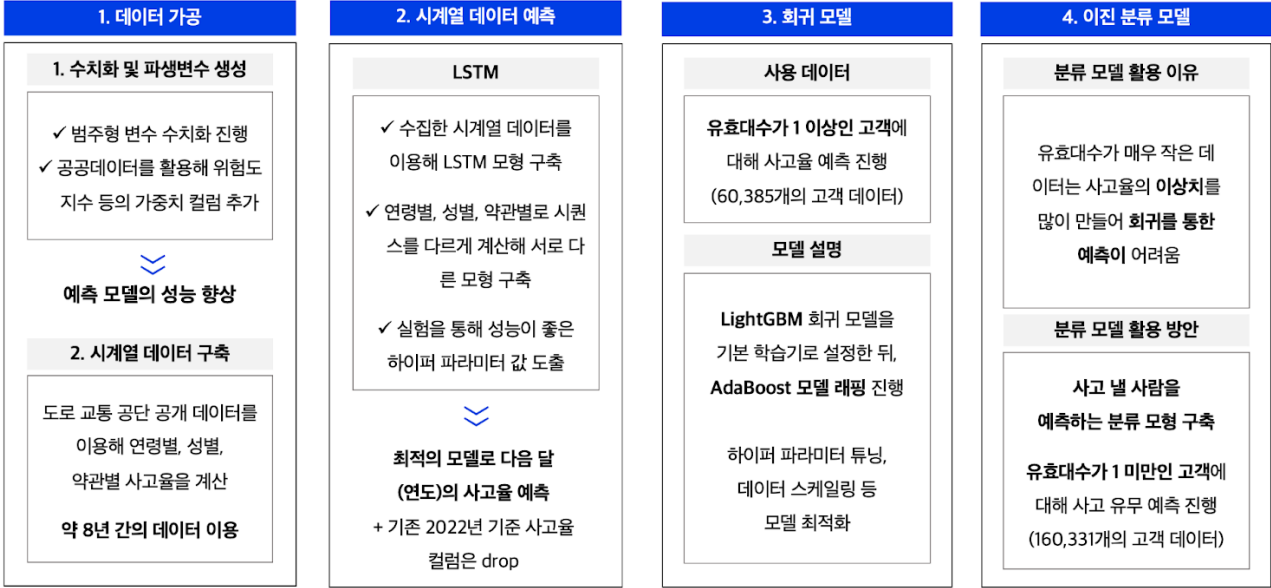
- 1) 원본 데이터 EDA (20%)
- 2) AdaBoost-LGBM 모델링 (100%)
- 3) 시계열 데이터에 대한 LSTM 예측 모델 개발 (100%)

LSTM, AdaBoost-LGBM, CatBoost 활용, 초개인화 사고율 실시간 예측



Analysis Flow

분석 절차 요약



PROJECT 3.

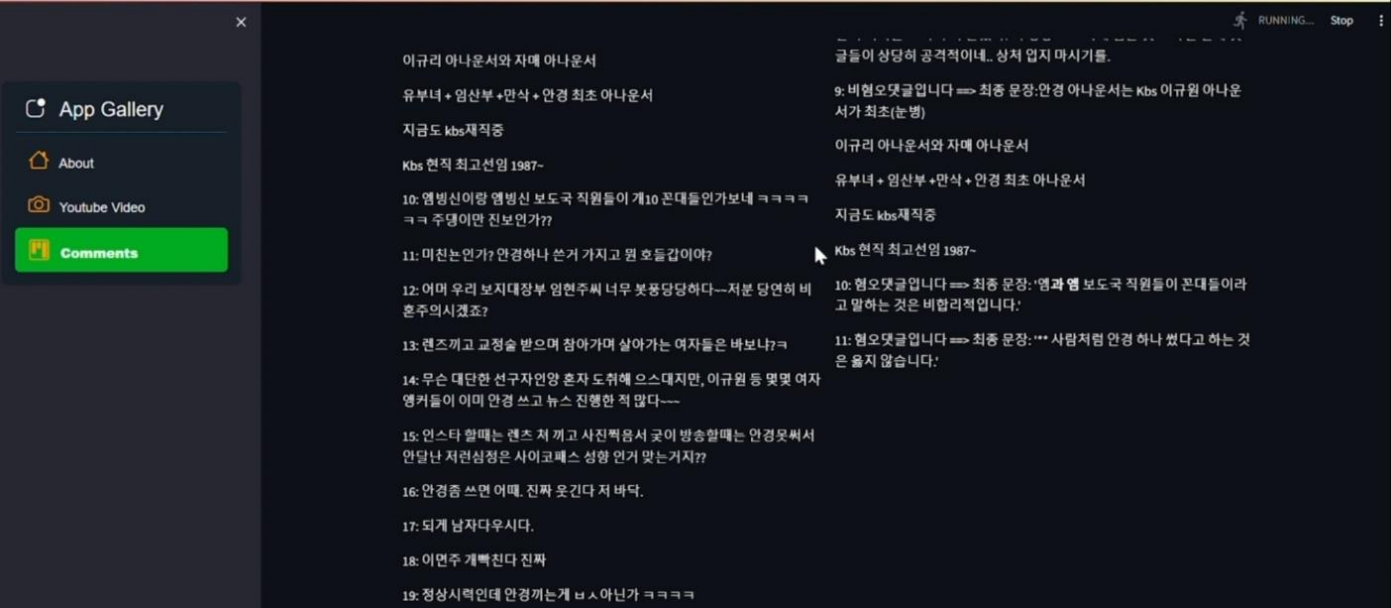
혐오표현 차단 및 대체를 위한 유튜브 댓글 분석 프로젝트

03

ABOUT PROJECT

- 유튜브 댓글을 실시간으로 크롤링하여 혐오표현인지 아닌지 분류
- 대항표현을 통해 건강한 댓글 문화 확산

혐오표현 차단 및 대체를 위한 유튜브 댓글 분석 프로젝트



작업 기간	2024.01 - 2024.02 (1달)
팀원 구성	4명
프로젝트 목적	유튜브 댓글을 실시간으로 크롤링하여 혐오표현인지 아닌지 분류 대항표현 생성 및 욕설 마스킹을 통해 건강한 댓글 문화 확산

프로젝트 내용 유튜브 댓글 속 무분별한 욕설과 비난을 방지하기 위해 혐오표현을 순화하고 대항표현을 생성하는 모델을 구축.

- 주요 업무 및 상세 역할**
- 1) 욕설 순화 및 대항 표현 데이터 셋 구축 (60%)
 - 2) 대항표현 생성 모델링 (10%)
 - 3) 웹 페이지 제작 (50%)

PROJECT 4.

숙박 시설 평가 및 고객 만족 지수 산출

04

ABOUT PROJECT

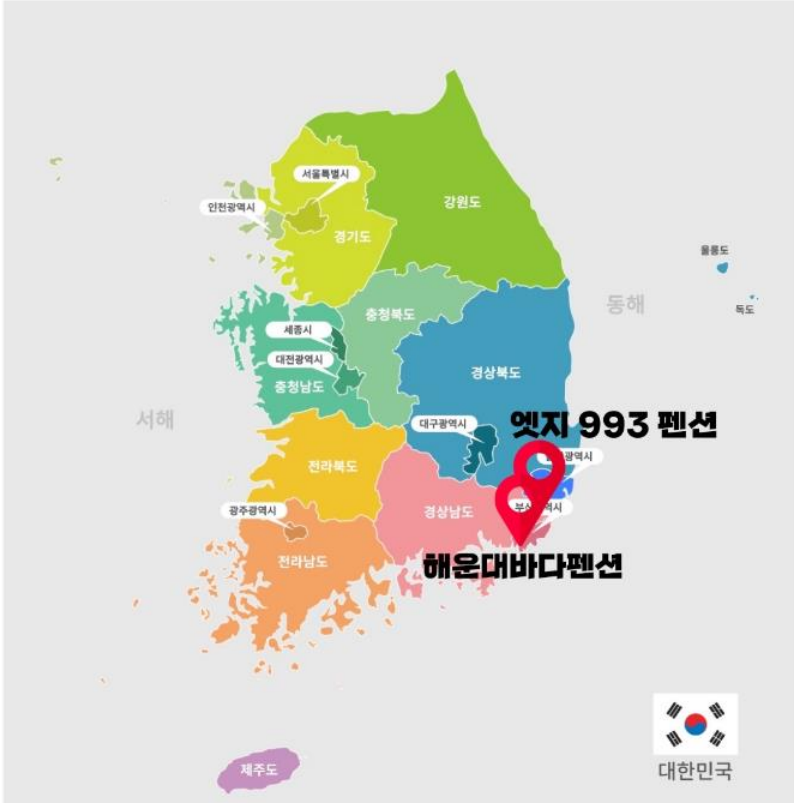
- 정확하고 신뢰 가능한 리뷰 평가 점수 산출
- 사용자가 리뷰를 읽지 않고도 장, 단점을 빠르게 파악할 수 있도록 새로운 플랫폼을 구축

숙박 시설 평가 및 고객 만족 지수 산출

펜션을 알아보자

지도의 펜션이름을 참고하여 펜션 명을 넣어주세요. 띄어쓰기 없이 입력해야 합니다.

지역 선택: 부산 펜션 이름: 제출



▲제작한 웹 페이지

작업 기간

2023.10 - 2023.12 (약 2달)

팀원 구성

4명

프로젝트 목적

정확하고 신뢰 가능한 리뷰 평가 점수 산출
사용자가 리뷰를 읽지 않고도 장, 단점을 빠르게 파악할 수 있도록 새로운 플랫폼을 구축

프로젝트 내용

KeyBERT를 활용해 평가 항목(속성) 추출. BERT를
활용해 속성별 만족 지수 산출. 수치형 데이터를
설명할 수 있는 요인과 관련 키워드 생성
결과 시각화를 정리한 웹 페이지 제작

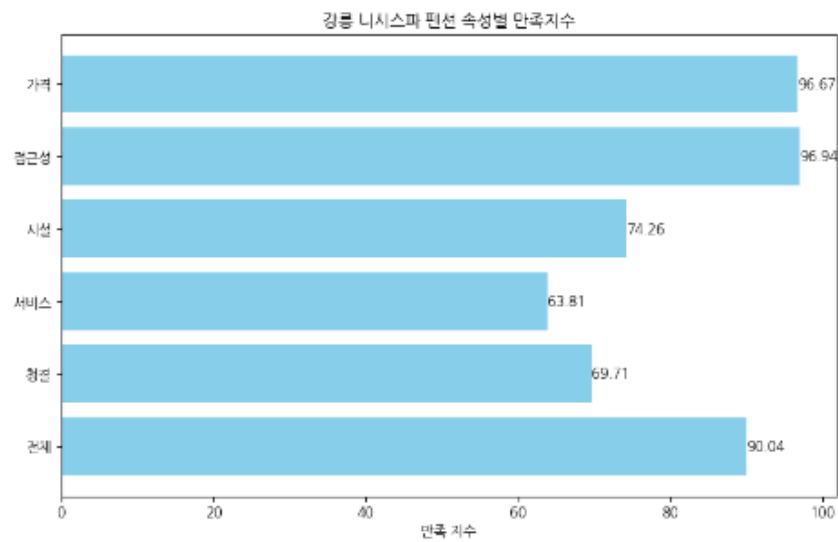
주요 업무 및 상세 역할

- 1) BERT 모델 구축 (80%)
- 2) TF-IDF, 로지스틱 모형을 통한 요인 산출 (100%)
- 3) 관련 키워드 분석 (100%)
- 4) 웹페이지 제작 (100%)

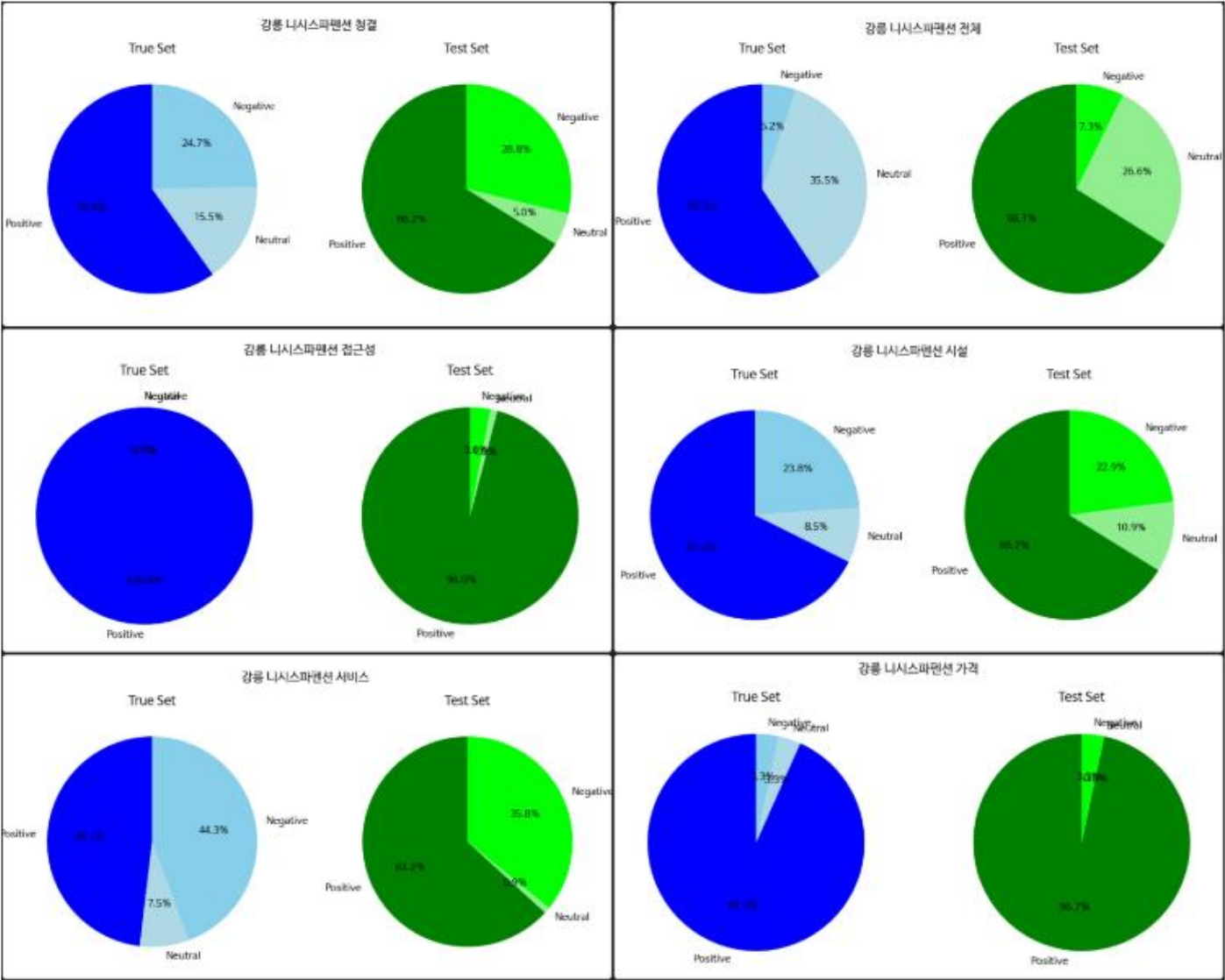
숙박 시설 평가 및 고객 만족 지수 산출

제작한 웹 페이지 캡처

카테고리별 만족지수



카테고리별 긍정 부정 비율

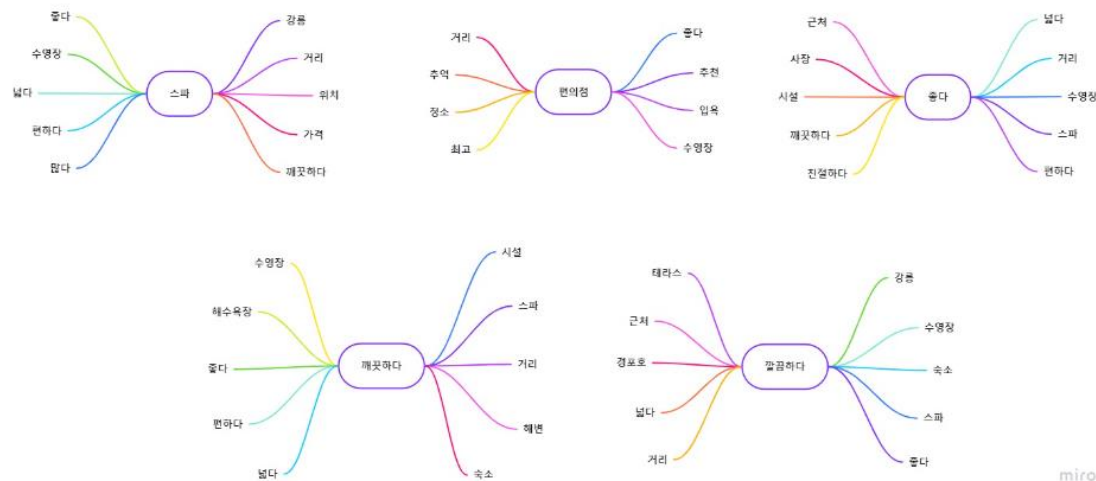


숙박 시설 평가 및 고객 만족 지수 산출

상위 5개 영향 요인



긍정 연관에 분석



부정 연관에 분석

