

## 데이터수집 보고서

### □ 개요

- 산출물 단계 : 모델배포
- 평가 산출물 : 개발된 LLM 연동 웹 애플리케이션
- 제출 일자 : 2025.04.02
- 깃허브 경로 : <https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN07-FINAL-2Team>

### 데이터 수집 목적

#### 데이터 수집의 필요성

##### 프로젝트 목표:

본 프로젝트는 사람 이미지만 보고 얼굴형을 분류하고 그 얼굴형에 맞는 안경 추천 및 그에 맞는 안경을 가상피팅 할 수 있는 모델을 개발 하는 것이 목표로 다양한 얼굴 이미지를 수집하여 모델 학습에 활용

##### 데이터의 중요성:

- 얼굴형 분류 모델의 정확도를 높이기 위해서는 다양한 얼굴형과 얼굴 이미지 데이터가 필요
- 각 얼굴형에 어울리는 안경 모델을 추천 및 피팅하려면 다양한 안경의 형태의 이미지 데이터가 필요

#### 데이터 활용 계획

##### 얼굴형 데이터

- **모델 훈련:** 수집한 얼굴형 데이터를 사용하여 CNN 모델 활용 후 데이터에 제일 적합한 모델을 찾아 전이학습 할 계획 (VGG16 모델 선정)

○모델 평가 및 개선: 더 정확한 얼굴형 인식을 위해  
데이터 전처리 과정에서 boxig 기법, 흑백 처리를  
해서 정확도를 높일 예정

#### 안경 데이터

수집한 안경 데이터를 사용하여 사용자가 가상피팅을 하는  
안경 데이터 활용 및 안경 이미지 제공에 활용

## 모델링 얼굴형 학습 데이터

### 얼굴형 이미지 데이터

- 목적: 정확한 얼굴형 분류를 위해 5가지 (Oval, Oblong, Square, Round, Heart) 데이터 수집
- 방법: kaggle에서 데이터셋 다운로드 (공개 데이터셋)



## 피팅 및 정보제공을 위한 안경 데이터

젠틀몬스터 공식 사이트 (<https://www.gentlemonster.com/kr/ko>)

BlueElephant 공식 사이트 (<https://blueelephant.co>)

### 안경 이미지 데이터용

- 목적: 안경피팅 및 안경 이미지 제공을 위해 수집
- 방법: 안경 이미지 캡처 및 누끼 수작업 (자체 데이터 수집)

### 안경 상세정보 데이터

- 목적: 사용자에게 안경 상세정보 제공을 위한 수집
- 방법: 자체 데이터 수집





수집 데이터(요약)

데이터 규모

총 데이터:

- 얼굴형데이터/ train/ 이미지 데이터(.jpg) 4000개  
(Heart, Round, Oval, Oblong, Square 각각 800개)
- 얼굴형데이터/ test/ 이미지 데이터(.jpg) 1000개  
(Heart, Round, Oval, Oblong, Square 각각 200개)
- 안경데이터/ 이미지 데이터(.png) 137개
- 안경데이터/ 엑셀(세부정보) 데이터

데이터 문제점 및 해결방안

문제점: 얼굴형 구분이 애매한 경우 발생

해결: 2명 이상의 검수자가 확인하여 다수결로 결정

데이터 저장

MariaDB 활용

선정 이유

- 안정성이 높음
- 데이터 일관성 유지
- 구조화된 데이터를 저장하고 빠르게 조회
- 향후 분석 및 활용을 위한 데이터 접근성 보장