

# 프로젝트형 일경험 결과보고서

☐ 프로젝트 유형 : 청년주도 프로젝트

☐ 프로젝트명 :

PredicTube: AI 기반 유튜브 영상 성과 예측 및 특성 최적화 서비스

☐ 팀명 : 벼락코딩

☐ 팀 구성원 : 임한빈, 이도현(중도포기), 정정훈

☐ 수행기간 : 9 월 25 일 ~ 2 월 1 일

☐ 프로젝트 개요

- 유튜브 업로드 전 데이터(제목, 썸네일 등)를 바탕으로 유튜브 성과(조회수)를 예측해주는 AI를 비롯해, 축적된 유튜브 영상 데이터셋을 활용해 다양한 관점에서의 분석 데이터를 제공하는 웹서비스 개발.

☐ 추진배경 및 목적

■ 유튜브의 대중성:

2022년 9월 현재, 대한민국 인구의 약 81%, 즉 4183만 명이 한 달 동안 13억 8천만 시간을 유튜브로 소비하고 있음.(통계: KOSIS 2022) 이는 유튜브가 단순한 동영상 플랫폼을 넘어서 광고와 마케팅의 핵심 플랫폼으로 자리매김하였음을 시사.

■ 유튜브의 경쟁 환경:

그러나 유튜브의 급격한 성장에 따라 수요보다 공급이 폭발적으로 늘어, 수많은 크리에이터들이 제한된 시청 시간 속에서 경쟁하게 됨. 이 때문에 영상의 제목과 썸네일은 크리에이터에게 시청자를 끌어들이는 핵심 요소가 되었음.

■ 제목 및 썸네일 선정의 어려움:

현재, 제목과 썸네일 선정에 관한 명확하고 객관적인 기준이 부족하여 크리에이터들은 이에 대해 많은 시간과 노력을 들이게 됨. 콘텐츠 제작이나 마케팅에 소비되는 시간과 비용이 증가한다는 것은 곧 생산성 저하와 경쟁력 약화로 이어져 디지털 미디어 콘텐츠 생태계의 발전이 저해될 것.

■ 목표 대상층

- 유튜브 크리에이터: 크리에이터의 눈높이에서 제목 및 썸네일 최적화를 도와, 채널 성장과 수익 증대를 목표로 하는 크리에이터 전체.
- 영상 마케팅 담당자 및 기업 브랜드: 유튜브를 활용한 마케팅 활동을 진행하는 기업 또는 개인.

□ 프로젝트 수행방법

- 개인별 역할

👉 팀장 임한빈

- 프로젝트 기획, 설계 및 총괄
- 프로젝트 관련 서류 작성 및 개발 일정 관리(PM)
- 데이터 수집 프로세스 자동화
- 데이터 전처리 및 검수
- AI 모델링, 학습, 성능평가 및 모델 최적화
- 웹 프론트엔드 및 백엔드
  - 메인, 결과, 시각화, 키워드 페이지 디자인 및 개발
  - 유튜브 채널ID를 등록해 구독자수를 가져오는 기능 구현
  - 히스토리 모달창 구현
  - 서비스 MVC 패턴 설계 및 개발
  - Google Cloud Vision API, chatGPT API 연결
  - Flask로부터 받는 데이터 가공
  - 결과 페이지 각 컴포넌트에 값 연결 및 세부 디자인 조정
  - 썸네일 Drag & Drop 기능 구현
  - 시각화 자세히/간단히 보기 토글 기능 구현
  - 각 페이지 카테고리 동기화 기능 구현
  - 자체 알파 테스트 후 각종 버그 픽스 및 UI·UX 개선
- 데이터베이스 테이블 설계 및 쿼리 작성

☞ 팀원 이도현 (중도포기)

· 웹 프론트엔드 및 백엔드

- 부트스트랩 기반 웹 페이지 프로토타입 개발
- 결과 페이지 레이아웃, 히스토리 모달창 디자인
- 도넛 차트 구현
- 구글 소셜 로그인 기능 구현
- Flask와 Tomcat 간의 서버 연결

☞ 팀원 정정훈

· 데이터 관리 및 시각화

- 추진 일정

- 1차 : 프로젝트 개선 사항 및 방향성 탐색 (모바일 앱 개발)
- 2차 : 모바일 앱 개발 기술 학습 (Flutter, Firebase 등)
- 3차 : 멘토링 후 개발 방향 조정 (모바일 앱 개발 중단, 웹 서비스 개선)
- 4차 : 웹 개선 사항 선정 및 웹 디자인 전면 수정
- 5차 : 웹 개발 및 추가 기능 구현 (채널ID 등록, chatGPT API 등)

□ 프로젝트 수행내용

- 주요 기능

■ 제목과 썸네일 이미지를 분석해 영상의 성과를 예측

- 사용자가 유튜브 영상을 업로드하기 전에 제목과 썸네일, 구독자수, 카테고리 정보만 가지고 조회수 구간을 예측해 봄으로써 더 최적의 제목과 썸네일로 수정하거나 콘텐츠 자체적인 이슈 정도를 확인하여 영상 주제를 선정하는 데에 도움을 줌.

#### ■ chatGPT의 제목 추천

- 사용자가 입력한 제목 텍스트로부터 핵심 키워드를 추출하여 chatGPT API를 통해 유튜브 영상 제목을 세 가지 추천받음. 이를 통해 본 서비스에서 제공하는 데이터 분석을 온전히 활용하기 어려워하는 사용자에게 더 접근성 좋게 제목 최적화를 도움.

#### ■ 예측 시도한 입력값과 결과값을 회원별 DB에 저장해 히스토리 조회

- 사용자가 시도했던 입력값들을 비교해보고 더 나은 결과값을 도출한 입력값을 실제 업로드 시의 제목과 썸네일로 확정하도록 도움을 줌.

#### ■ 카테고리별, 변수별로 조회수와 상관관계를 그래프로 시각화

- 각 카테고리별 경향을 눈으로 확인하여 특성을 파악하고 유튜브 영상 업로드에 대한 전략을 세우는 데에 도움을 줌. 또한, 사용자의 입력값과 비교하여 어떤 방향으로 수정하면 좋을지 알 수 있게 함.

#### ■ 카테고리별 키워드 빈도수 랭킹

- 제목과 썸네일에 들어갈 키워드로 어떤 것이 좋을지에 대해 객관적인 데이터를 제공하여 키워드 선정에 도움을 줌. 또한, 각 분야별로 어떤 키워드가 많이 이슈되고 있는지를 확인할 수 있어 콘텐츠 소재 탐색에 도움을 줌.

#### - 적용 기술

##### ■ 인공지능

■ Python (Keras, Tensorflow) - Jupyter Notebook

개발환경 구축, 데이터 분석 및 시각화, 데이터 분류 및 전처리, 머신러닝 및 딥러닝, 자연어처리

##### ■ 웹 서비스

■ JSP (HTML5, CSS3, Java), JavaScript, Tomcat, Flask (Python)

■ DB (MySQL)

## - 개발 내용

### (1) 데이터 수집 및 전처리

#### ○ YouTube Data API v3 활용:

☞ 대량의 유튜브 영상 데이터(약 10만 건)를 자동으로 수집함. 이를 통해 영상의 제목, 썸네일 URL, 조회 수, 좋아요 수, 댓글 수 등의 기본적인 유튜브 데이터를 획득함.

#### ○ Google Cloud Vision API 활용:

☞ OCR 기능: 썸네일 이미지 내의 텍스트 정보를 추출하고, 텍스트의 크기와 길이를 분석하여 데이터화함.

☞ 얼굴 인식 기능: 썸네일 내에 사람의 얼굴 유무를 판별함.

☞ 유해성 검사 기능: 썸네일의 폭력적이거나 선정적인 요소를 판별하고, 그에 따른 유해성 수치를 도출함.

☞ 데이터 저장 및 전처리: 수집된 모든 데이터를 중앙 데이터베이스에 저장함. 이후, 필요한 형태로 데이터 전처리를 진행하여 AI 모델 학습에 바로 사용할 수 있도록 준비함. (불용어 제거, 토큰화, 패딩, 임베딩, 스케일링 등)

### (2) AI 모델링 및 훈련

#### ○ 자연어처리 (NLP, Natural Language Processing)

☞ LSTM (Long Short-Term Memory) 모델을 활용하여 제목과 썸네일 텍스트를 처리함.

#### ○ 수치 특성 분석

☞ MLP (Multi-Layer Perceptron) 모델을 활용하여 구독자 수, 제목 길이, 얼굴 유무, 유해성 수치 등의 정량적 수치 특성을 처리함.

- 다중입력 다중출력 (Multi-Input Multi-Output) 모델

- ☞ LSTM과 MLP를 연결하여 MIMO 모델을 생성함.

- 모델 훈련 및 성능 평가:

- ☞ 준비된 데이터셋을 사용하여 모델을 훈련시킴. 각 카테고리별 모델들의 성능을 시각적으로 확인하고 비교·분석함. 최적의 결과를 얻기 위해 hyperparameter tuning 등의 기법을 활용함.

### (3) 웹플랫폼 개발

- 구글 소셜 로그인 기능:

- ☞ 사용자는 구글 계정으로 간편하게 로그인할 수 있음.

- 유튜브 채널 ID 등록:

- ☞ 유튜브 채널 ID를 사용자 계정에 대해 등록하면 이후 일일이 채널 ID나 구독자 수를 입력하는 번거로움 없이 자동으로 구독자 수를 가져와서 예측에 활용함.

- 조회수 예측:

- ☞ 사용자가 영상 제목과 썸네일 이미지 파일을 업로드하면, AI 모델이 예측한 영상의 성과를 보여줌.

- chatGPT의 제목 추천:

- ☞ chatGPT API를 통해 사용자가 입력한 제목의 핵심 키워드를 바탕으로 유튜브 영상 제목을 세 가지 추천해줌.

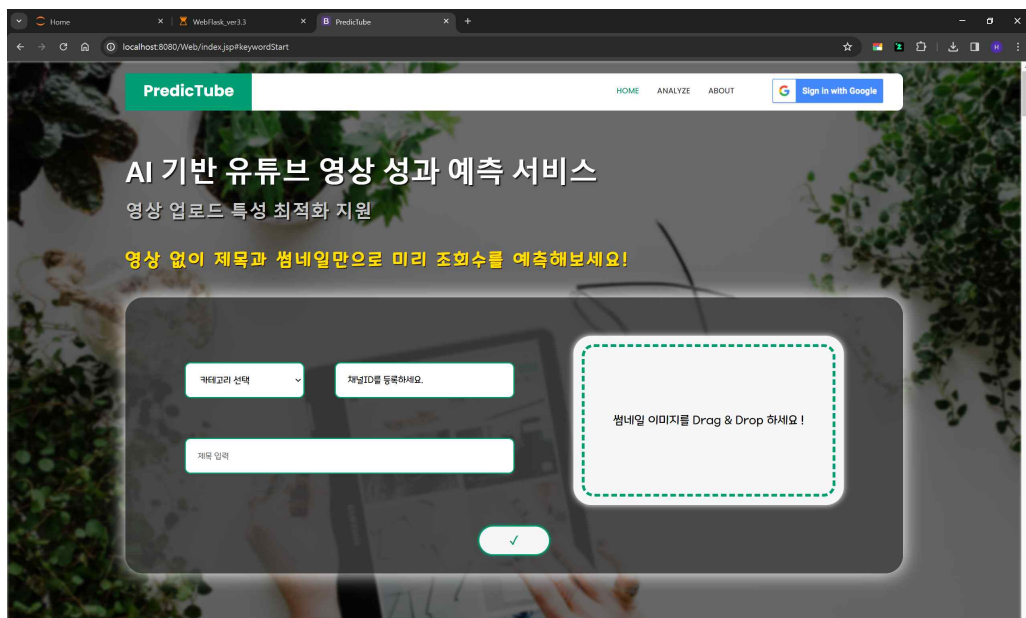
- 데이터 시각화:

- ☞ 영상의 성과와 각 변수(썸네일 텍스트 크기, 길이, 얼굴 유무 등) 간의 상관관계를 그래프로 보여줌. 이를 통해 사용자는 어떤 변수가 영상 성과에 큰 영향을 미치는지 한 눈에 파악할 수 있음.

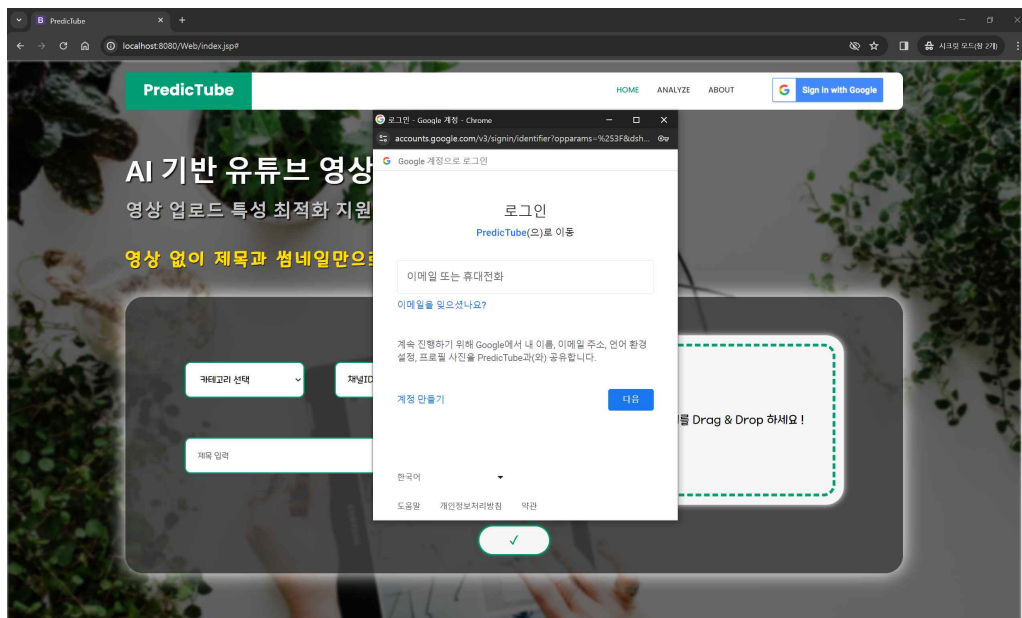
○ 키워드 랭킹 제시:

- ☞ 모델 학습에 사용된 데이터셋을 바탕으로 카테고리별로 키워드 빈도수 랭킹을 제시한다. 이로써 사용자의 키워드 선정을 수월하게 함.

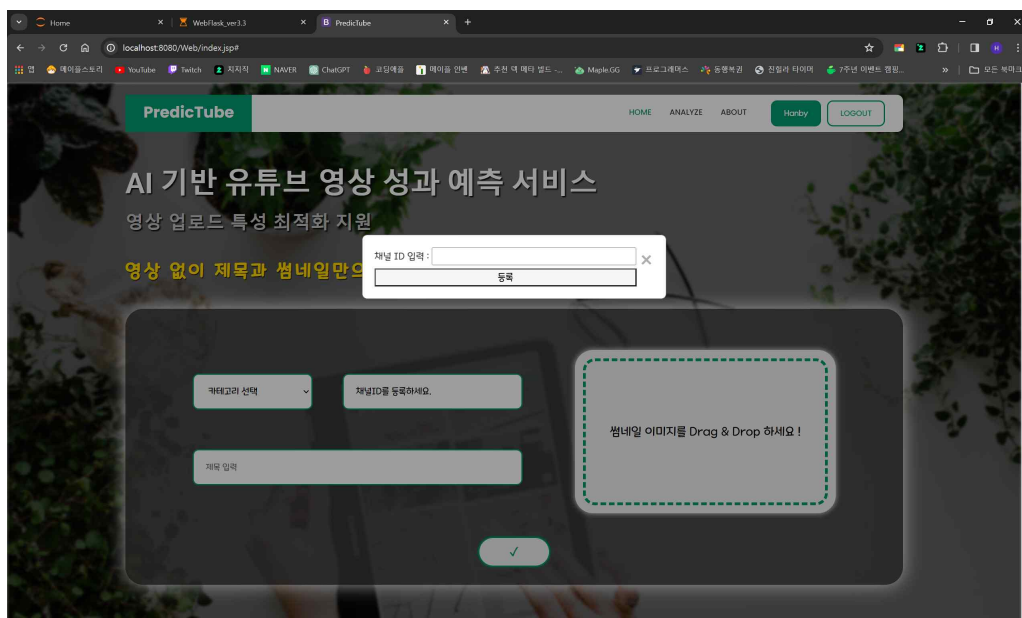
□ 프로젝트 결과물



< 메인 페이지 >

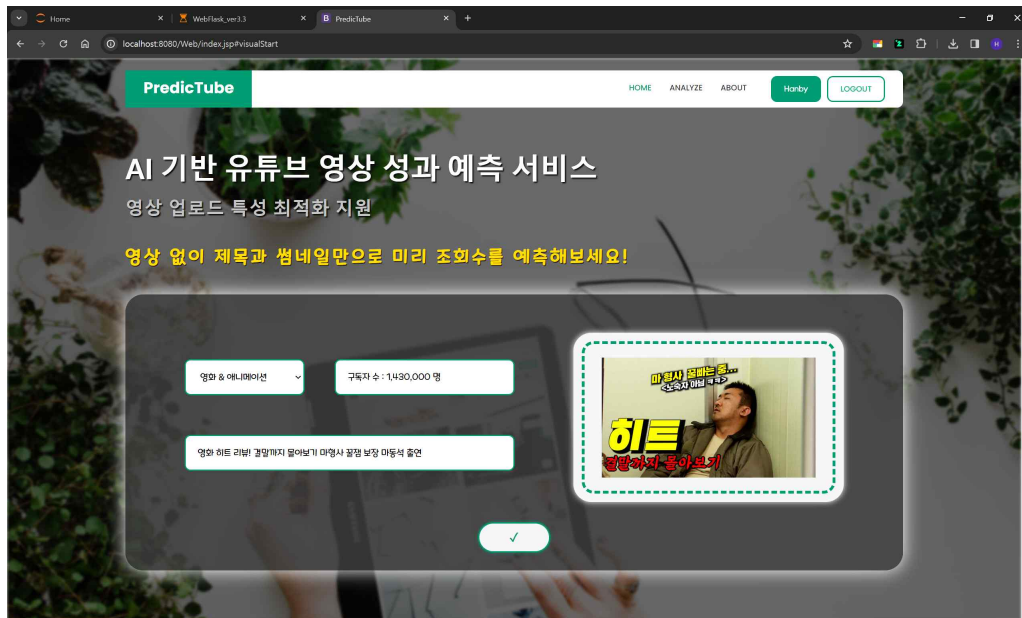


< 구글 로그인 >

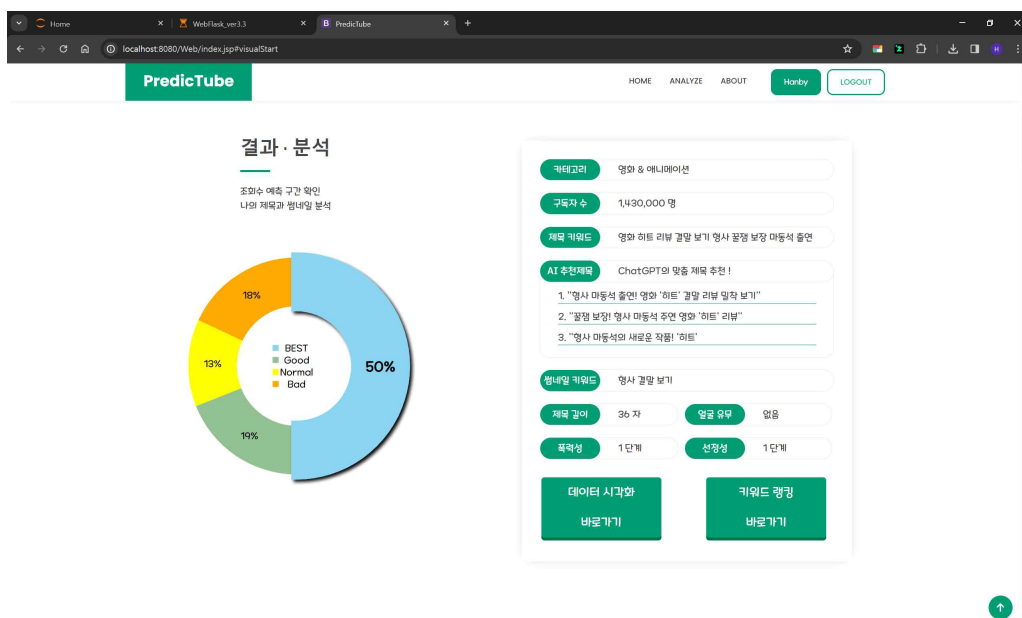


< 유튜브 채널 ID 등록 모달창 >

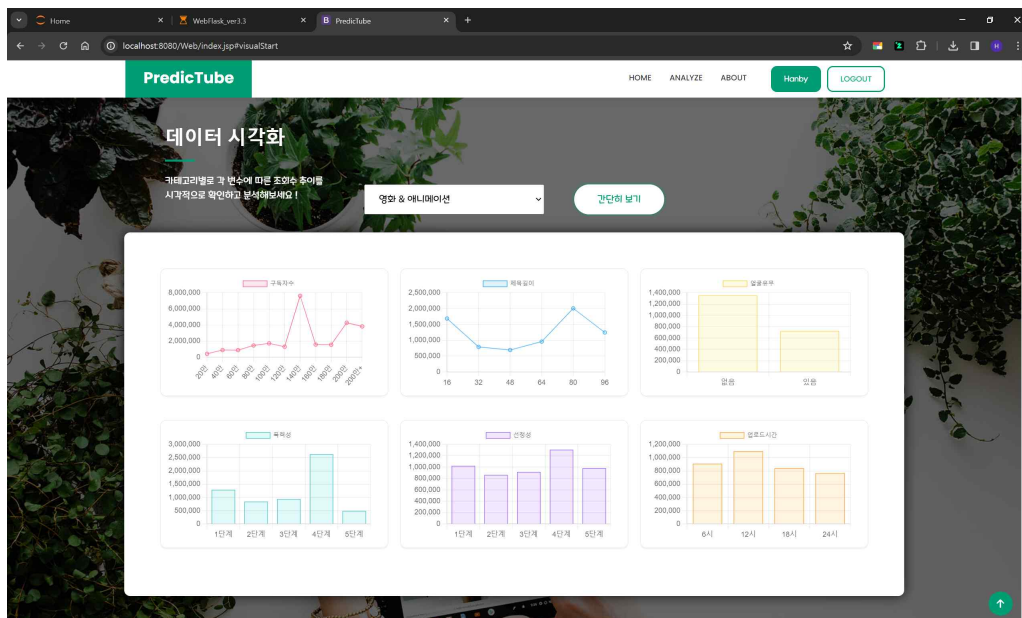




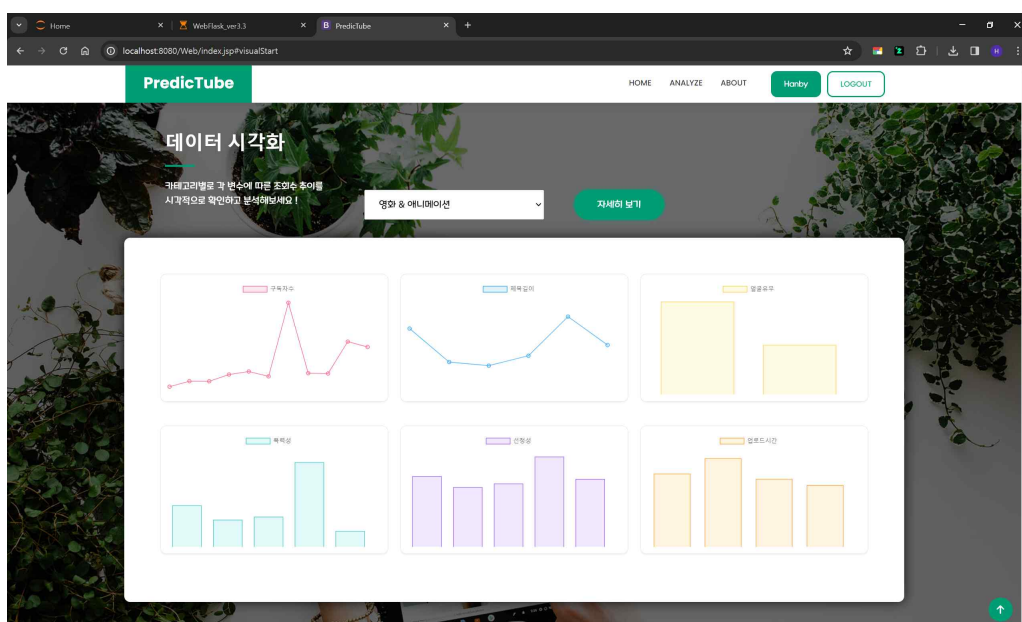
< Drag & Drop 으로 썸네일 이미지 업로드 >



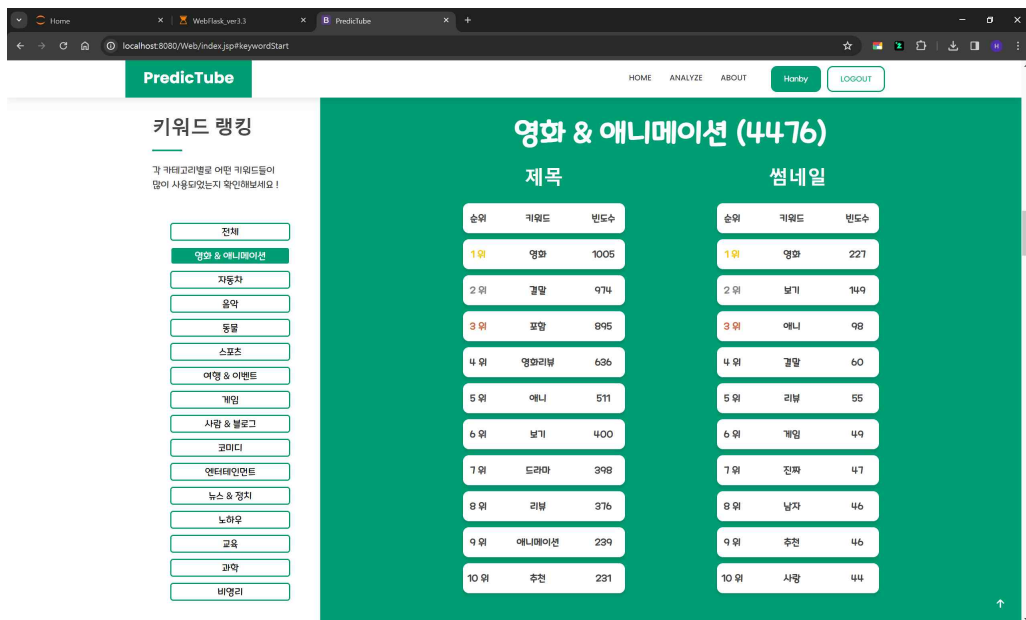
< 결과 페이지 >



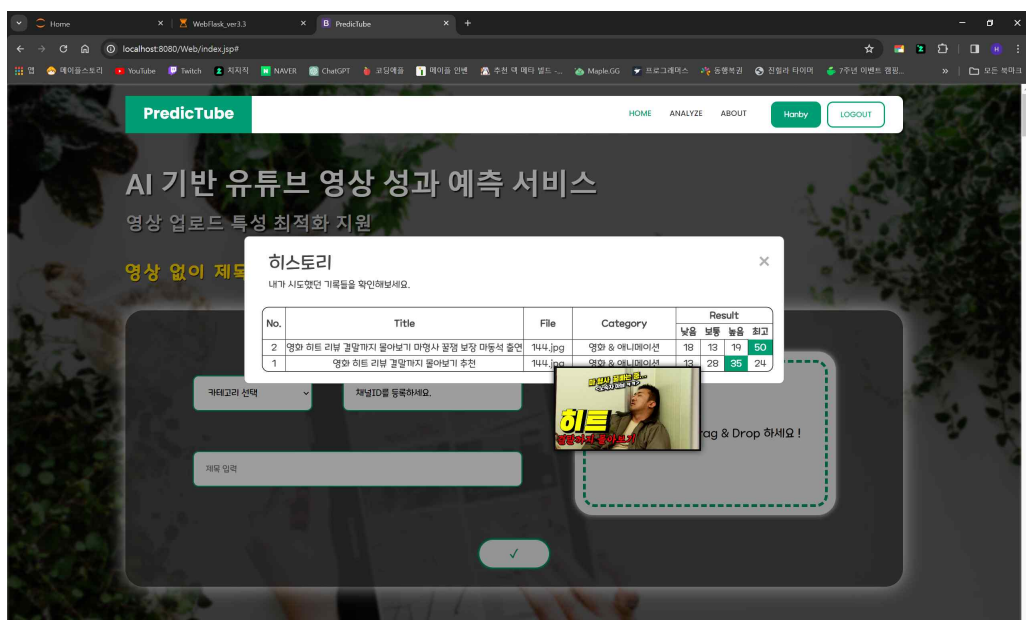
< 데이터 시각화 페이지 – 자세히 보기 >



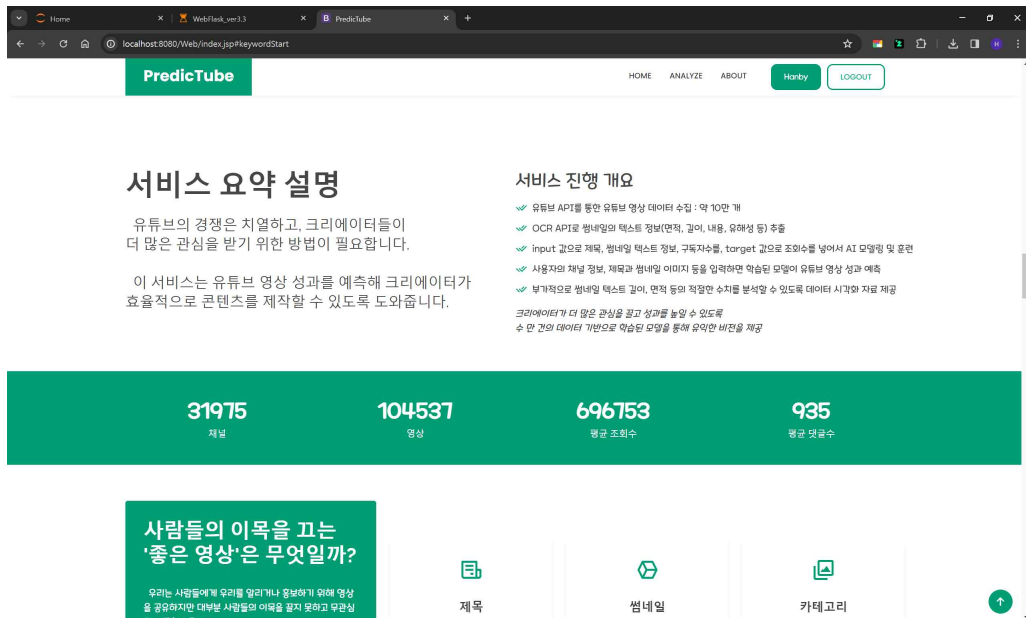
< 데이터 시각화 페이지 – 간단히 보기 >



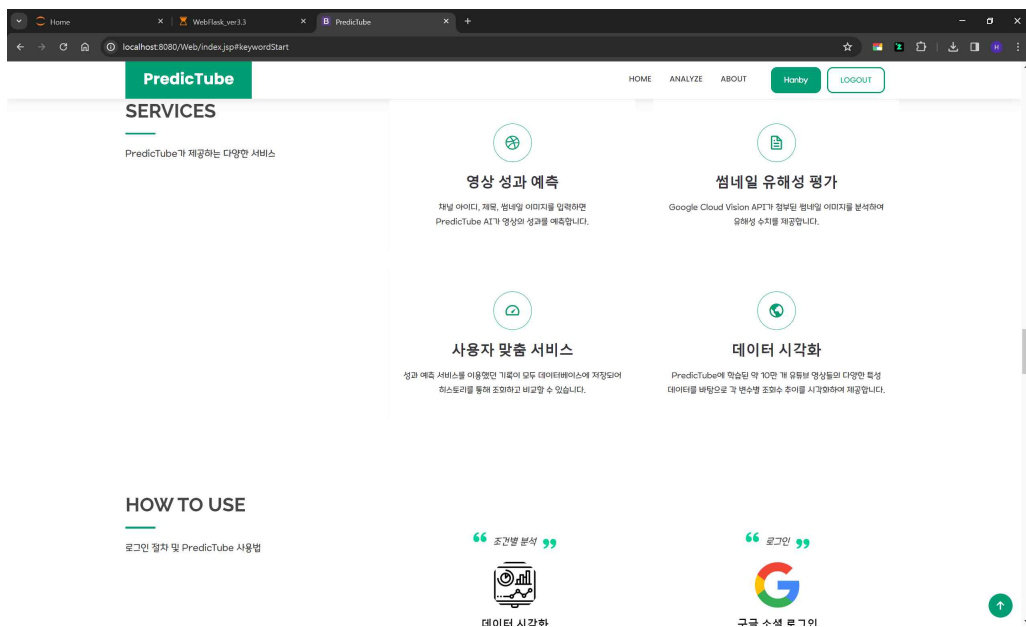
< 키워드 랭킹 페이지 >



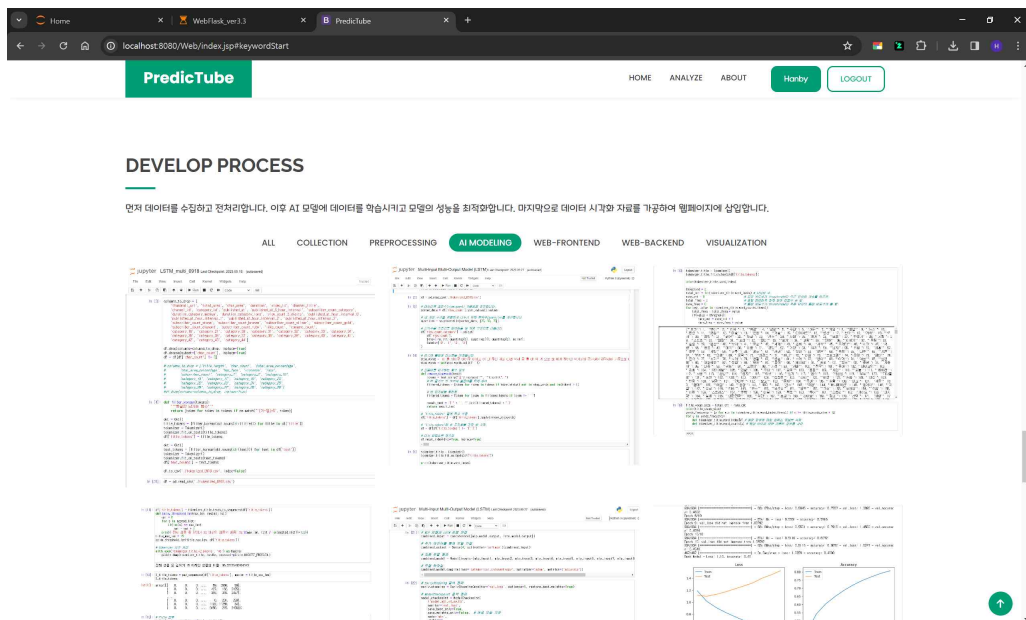
< 히스토리 모달창 및 썸네일 미리보기 >



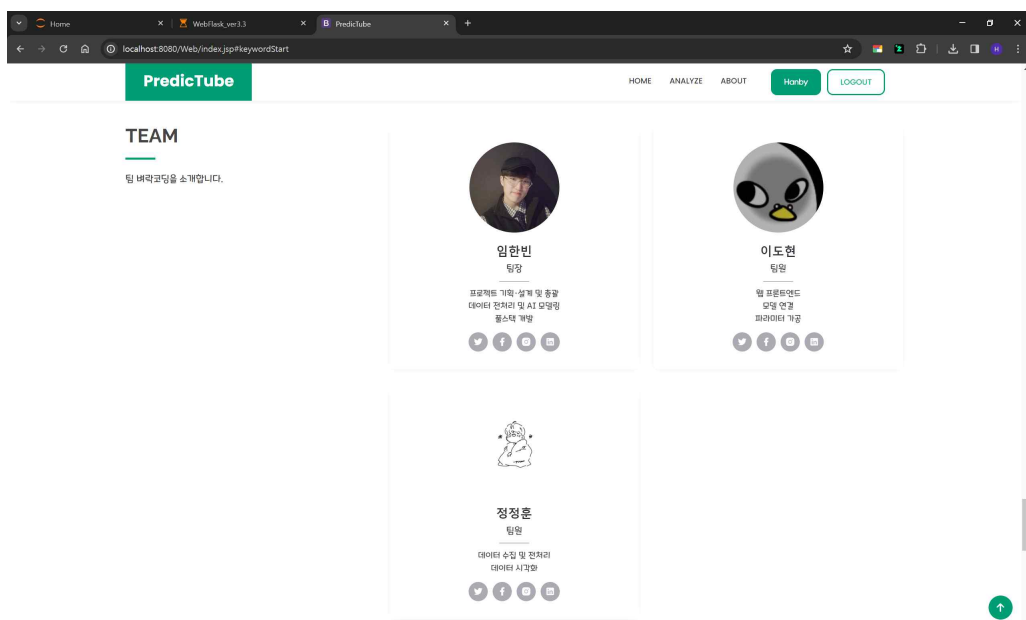
< 서비스 설명 페이지 >



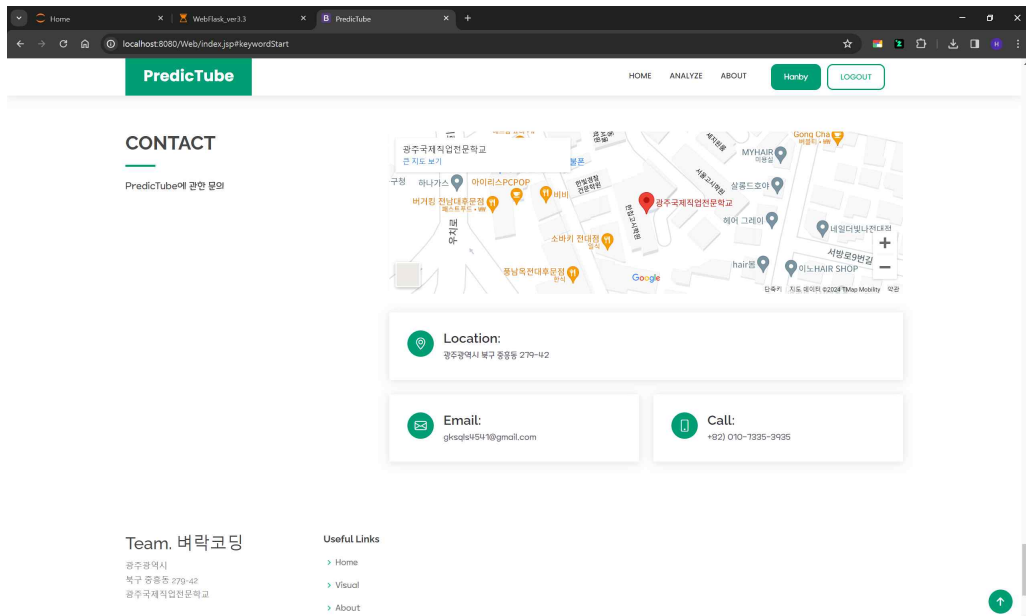
< 서비스 정보 및 사용법 페이지 >



< 개발 과정 페이지 >



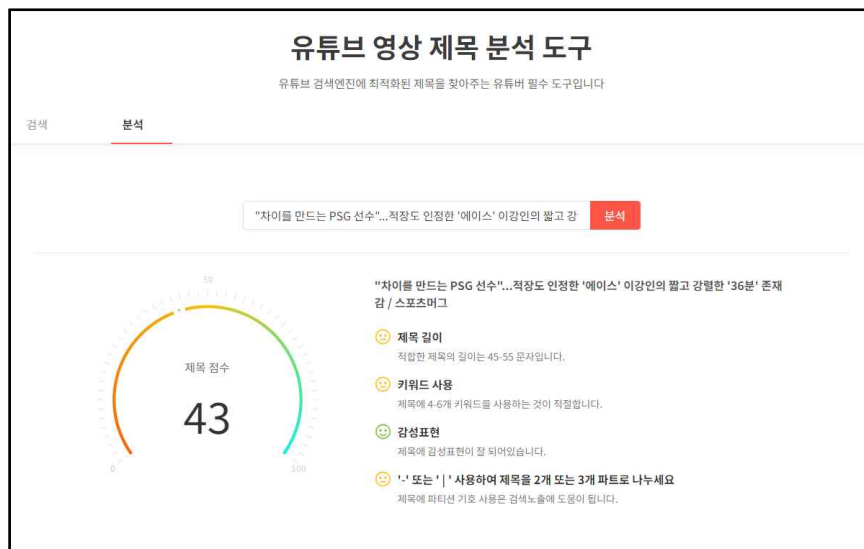
< 개발팀 소개 페이지 >



< 버그 제보 및 문의 페이지 >

## □ 고찰 및 기대효과

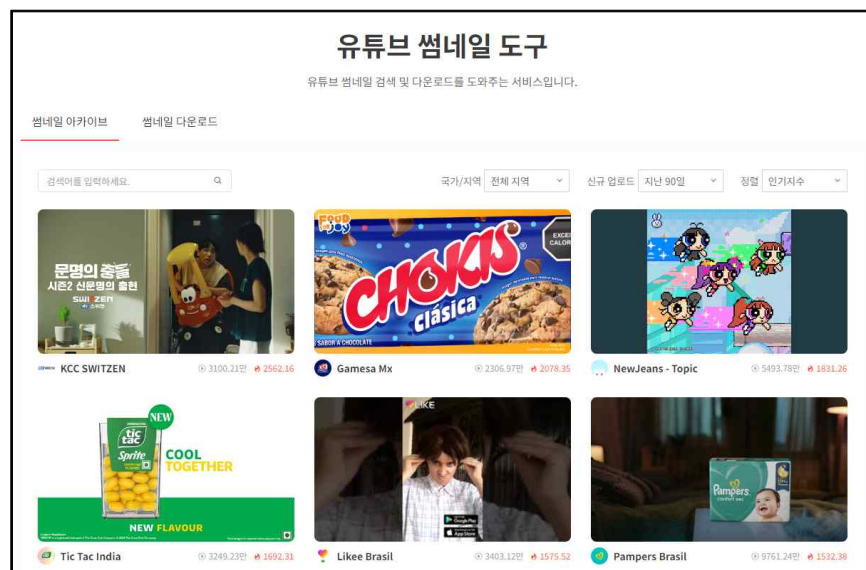
- 기존 경쟁 서비스 분석 및 차별점 - '녹스인플루언서'



'녹스인플루언서' - 유튜브 영상 제목 분석 도구

유튜브 제목을 제시했을 때, 해당 제목에 대해 '제목 길이', '키워드 개수', '감성표현', '파티션 기호' 등을 분석 요소로 사용하여 항목별 평가 및 제목 점수 출력.

→ AI 기반 제목 분석 X, 개발진의 주관적인 기준 제시



'뉴스인플루언서' - 유튜브 썸네일 도구

기존에 업로드된 유튜브 영상 썸네일 조회 및 다운로드.

→ 썸네일 분석 X



**키워드 검색 시** '시간대별 트렌드 지수', '관련 뉴스', '관련 인기 동영상' 출력.

→ 키워드 추천 / 키워드 랭킹 기능 X

- 기대효과 및 활용 방안

■ 객관적인 분석 기반의 제안:

크리에이터들에게 제목과 썸네일의 최적화에 대한 명확한 지표와 방향을 제시, 이를 통해 크리에이터의 작업 효율성 및 퀄리티 향상을 기대할 수 있음.

## ■ 유튜브 영상 마케팅 최적화:

제공되는 데이터 및 분석을 통해, 크리에이터뿐만 아니라 마케팅 담당자나 기업 브랜드 등도 유튜브 콘텐츠의 효과적인 마케팅 전략을 세울 수 있을 것.

■ 심리 및 행동 분석 연구:

연구 기관들은 다양한 데이터에 접근하여 디지털 미디어와 소비자 행동에 대한 연구를 실시하고 새로운 인사이트를 얻을 수 있을 것.



#### - 프로젝트를 진행하며 느낀 점

평소에 이용하던 웹사이트들이 이렇게 많은 시간과 노력을 투자해야 만들어진다는 것을 알게 되었고, 프로젝트 주제 자체가 굉장히 확장성이 넓어서 개발의 방향성이나 우선 순위를 정하기 어려웠음. 웹서비스 속 기능에 대해서 더 다양하고 참신한 아이디어들이 많았으나 시간과 기술력의 부족으로 모두 구현하지 못했다는 점이 아쉬웠음. 프로젝트를 설계함에 있어서 개발력을 가늠하고 개발 범위를 설정하는 노하우를 경험으로 알게 되었음. 팀원들과의 원활한 소통의 중요성을 절실히 느꼈고, 지속적인 피드백을 바탕으로 본래 기획·설계한 수준 이상까지 수정해 나가며 퀄리티를 올리는 것이 프로젝트의 완성도를 높이는 데에 필수적이라는 것을 깨달았음.

#### □ 향후계획

- 로컬 환경에서만 개발했기 때문에 외부 서버나 DB를 활용해서 서비스를 배포하고 베타테스트를 진행하여 프로젝트의 완성도를 높이고 추후 정식 서비스 출시 가능성 검토. 제공하는 AI의 모델 버전, 분석 데이터의 종류나 양 등에 차별점을 두어 유료 서비스로서 사업화의 여지가 있다고 판단됨.

#### □ 참고 문헌

##### **딥 러닝을 이용한 자연어 처리**

- <https://wikidocs.net/book/2155>

##### **YouTube Data API**

- <https://developers.google.com/youtube/v3/getting-started?hl=ko>

##### **Google Cloud Vision API**

- <https://cloud.google.com/vision/docs/features-list?hl=ko>