**문제해결 빅데이터 활용 프로젝트 기획안**

기획안 작성일자 : 2023.10.30

|  |  |
| --- | --- |
| 조 명 | 1조 : 구독해조 |
| 조 원 | 조장: 이서연 조원: 강민주 김도형 김현정 박준범 |
| 프로젝트 주제 및 개요 | 유튜브 인기 급상승 동영상 및 세부 카테고리 분석 |
| 프로젝트 수행 방향 | ● 분석 방향  - 카테고리 별 업로드 시간 및 요일, 관심 키워드 분석을 통해  기업, 유튜브 초심자, 이용자에게 필요한 정보 제공  ● 활용 데이터  - Kaggle, YouTube, 자막 추출 데이터, 음성 데이터 추출  ● 분석 도구 및 기술(모델)  pydub, speech\_recognition, pytube, bs4(BeautifulSoup), youtubeAPI, moviepy, googleAPI, konlpy, RandForest, KoBERT, LDA, Tf-Idf벡터라이즈 등… |
| 프로젝트 조직  (구성원 및 역할) | ● 역할분담  조장: 이서연  -**자막 추출 담당(2명):** 이서연 김현정  -**자막 정제** **담당(2명):** 강민주 박준범  - **음성 데이터 추출 코드 작성**: 김도형  -Human Labeling을 통해 **데이터 추출(5명)**  -**전체 데이터 경향성 분석(3명)**: 김도형 이서연  -**소주제별(전체 분석 이후) 시각화 및 모델링** – 강민주 김현정 박준범  **리포트(1명)**- 이서연  **ppt(1명)**-김현정  **발표(1명)**-강민주 |
| 프로젝트 추진 일정 | ● 주요 일정  10/06~ 10/12: 프로젝트 수행계획 수립 및 주제 탐색 리서치  10/12 ~ 10/21: 데이터 추출 코드 작성 및 추출 및 전처리  10/17 ~ 10/21: 카테고리 전체 경향성 시각화 및 모델링 기반 분석  10/21 ~ 10/24: 각 카테고리별 시각화 및 모델링  10/22 ~ 10/28: 보고서 및 ppt 제작, 발표 준비  10/29~ 10/31: 발표 준비 마무리 |