**비정형 데이터 활용 분석 프로젝트 기획안**

**2024년 5월 22일**

**과정명: 금융&마케팅 데이터 분석(Python)**

| 프로젝트 주제 | SNS 댓글 텍스트 분석을 통한 Samsung galaxy s24 제품의 강점과 약점 분석 |
| --- | --- |
| 팀 명 | 땅콩버터 |
| 팀 원 | 김준홍, 박진수, 박현진, 장준하 |
| 프로젝트 내용  해결하고자 하는 문제  최종 산출물의 청사진 | [고객과의 소통을 위한 여론 분석]  SNS(유튜브) 댓글 텍스트 감성분석을 통해 **Samsung galaxy s24** 제품에 대한 소비자의 의견을 알아보고자 한다. 이를 통해 제품에 대해 소비자가 느끼는 제품의 강점과 문제점을 도출하는 목표를 가지고 프로젝트를 수행한다.  최종 산출물: 긍정적 댓글, 부정적 댓글 각각 빈도수가 높은 키워드 시각화 |
| 프로젝트 추진 일정 | 기획(3d) ~5/22  수집(2d) ~5/22  전처리(2d) ~5/24  데이터 탐색(EDA) (2d) ~5/28  분석(2d) ~5/28  검토(1d) ~5/29  결과 도출 ~5/30  발표 ~5/31 |
| 프로젝트 수행 방향  수행 방법/도구 | 1. 수집   * sns 데이터 댓글 크롤링 - selenium   2. 전처리   * 결측치 및 중복 데이터 제거 - pandas * 긍정, 부정 문장 분류 - hugging face(pipeline)   3. 데이터 탐색(EDA)   * 데이터 시각화 및 비율 확인 - seaborn, matplotlib, wordcloud   4. 분석   * 키워드 간 연관성 분석 * 빈도수가 상위에 있는 키워드 추출 - collections   5. 결과 발표   * 분석을 통해 산출 된 키워드를 바탕으로 제품의 강점과 약점 파악 |