

# 프로젝트 헌장\_4팀

## 1. 프로젝트 개요

**프로젝트명:** 감성 기반 소셜 미디어 트렌드 예측 모델 개발

**수행 과목:** 빅데이터 이해와 분석

**프로젝트 기간:** 2025.03.17 ~ 2025.06.10

**팀 구성:** 현종민(팀장), 신예원(팀원), 김채은(팀원)

## 2. 프로젝트 배경 및 목적

**배경:**

- 소셜 미디어는 소비자의 의견과 감성이 실시간으로 표출되는 공간으로, 시장 트렌드를 선행적으로 반영하는 경향이 있음
- 기존의 트렌드 예측 방법은 과거 데이터에 의존하여 급변하는 시장 환경에 대응하기 어려움
- 소셜 미디어 데이터의 감성 분석을 통해 보다 선제적인 트렌드 예측이 가능할 것으로 기대됨

**목적:**

- 소셜 미디어에서 표현되는 대중의 감성과 의견을 분석하여 미래 트렌드나 매출과 같은 비즈니스 지표를 예측하는 모델 개발
- 텍스트 마이닝과 딥러닝 기술을 활용한 감성 분석 시스템 구축
- 실제 비즈니스 지표와 소셜 미디어 감성 데이터 간의 상관관계 검증

**기대효과:**

- 소비자의 감성 변화를 조기에 감지하여 기업의 의사결정 지원
- 마케팅 및 제품 개발 전략 수립에 활용 가능한 예측 인사이트 제공
- 소셜 미디어 데이터와 비즈니스 성과 지표 간의 연관성 파악을 통한 새로운 분석 관점 제시

## 3. 프로젝트 범위

**분석 대상:**

- 국내 주요 소셜 미디어(트위터, 인스타그램, 레딧) 데이터
- 특정 브랜드 또는 제품 카테고리(전자기기, 패션, 식품 중 선정)에 관한 소셜 미디어 게시물
- 최근 6개월간의 소셜 미디어 데이터와 해당 기간의 실제 판매 데이터

**주요 분석 내용:**

- 소셜 미디어 텍스트 데이터의 감성 분석(긍정/중립/부정)
- 감성 트렌드와 비즈니스 지표(판매량, 주가 등) 간의 상관관계 분석
- 감성 데이터를 활용한 시계열 예측 모델 개발 및 검증

**제외 사항:**

- 이미지, 동영상 등 비텍스트 데이터의 감성 분석
- 6개월 이상의 장기 트렌드 예측
- 소셜 미디어 이외의 온라인 채널(블로그, 뉴스 등) 분석

## 4. 데이터 수집 계획

- 데이터 소스: 트위터 API, 레딧 API
- 인스타그램 크롤링 데이터

- 분석 대상 브랜드/제품의 판매 데이터(공개 데이터셋 또는 가상 데이터)

#### 수집 방법:

- 소셜 미디어 API를 활용한 프로그래밍 방식의 데이터 수집
- Python 라이브러리(Tweepy, PRAW)를 활용한 소셜 미디어 데이터 수집
- 웹 크롤링 기법을 활용한 인스타그램 데이터 수집

#### 데이터 형태:

- 텍스트 데이터(게시물, 댓글, 해시태그)
- 메타데이터(작성 일시, 사용자 정보, 좋아요/공유 수 등)
- 시계열 형태의 판매 데이터 or 재무재표에서의 매출

#### 필요 변수:

- 텍스트 내용
- 게시 시간 및 날짜
- 인게이지먼트 지표(좋아요, 공유, 댓글 수)
- 해시태그
- 판매량 또는 관련 비즈니스 지표

## 5. 분석 방법론

#### 데이터 전처리:

- 텍스트 클렌징(특수문자 제거, 이모지 처리, URL 제거)
- 토큰화 및 불용어 제거
- 한국어 데이터의 경우 형태소 분석
- 시간별 데이터 집계 및 정규화

#### 탐색적 데이터 분석(EDA):

- 시간별 감성 분포 시각화
- 주요 키워드 빈도 분석
- 감성 트렌드와 판매 데이터의 시계열 비교 분석
- 인게이지먼트와 감성 간의 관계 분석

#### 분석 모델:

- BERT 기반 감성 분석 모델(KoBERT for 한국어)
- 감성-볼륨 복합 지표 개발
- LSTM 기반 시계열 예측 모델
- 다중 입력 변수를 활용한 트렌드 예측 모델

#### 평가 방법:

- 감성 분석 모델: 정확도, F1 스코어, 혼동 행렬
- 예측 모델: MAE(Mean Absolute Error), RMSE(Root Mean Squared Error),  $R^2$  점수
- 실제 비즈니스 지표와의 상관관계 분석을 통한 모델 유효성 검증

## 6. 필요 기술 및 도구

프로그래밍 언어: Python,

라이브러리/패키지:

- 데이터 처리: Pandas, NumPy

- 텍스트 처리: NLTK, SpaCy, KoNLPy(한국어)
- 딥러닝: TensorFlow, PyTorch, Transformers
- API 연동: Tweepy, PRAW
- 웹 크롤링: BeautifulSoup, Selenium

개발 환경: VS Code

## 7. 프로젝트 일정

단계	활동 내용	시작일	종료일	담당자
계획	프로젝트 계획 수립 및 데이터 소스 탐색	2025.03.17	2025.03.24	현종민
데이터 수집	API 연동 및 크롤링 시스템 구축	2025.03.25	2025.04.07	신예원
전처리	텍스트 정제 및 감성 분석 데이터셋 구축	2025.04.08	2025.04.21	김채은
모델 개발	BERT 기반 감성 분석 모델 구현	2025.04.22	2025.05.05	신예원
분석	감성-비즈니스 지표 상관관계 분석	2025.05.06	2025.05.12	김채은
예측 모델링	LSTM 기반 트렌드 예측 모델 개발	2025.05.13	2025.05.26	현종민
결과 도출	모델 평가 및 결과 해석	2025.05.27	2025.06.02	전체 팀원
보고서 작성	최종 보고서 작성	2025.06.03	2025.06.07	현종민
발표 준비	발표 자료 준비 및 리허설	2025.06.08	2025.06.10	전체 팀원