**기업요구사항 기반의 문제해결 프로젝트 기획안**

**2021년 08월 27일**

| **프로젝트 조** | **6조** | |
| --- | --- | --- |
| **프로젝트 팀원** | 팀명: 업빛투  팀장: (정) 정길종, (부) 김형림  팀원: (사이언스) 윤보람, 채길호 (엔지니어링) 인태우 | |
| **프로젝트 주제** | **뉴스 기사 및 유튜브 텍스트의 감성 분석과**  **딥러닝을 이용한 주가 등락 예측 서비스 구현** | |
| **프로젝트 목적** | **직장인들이 출퇴근 시간에 분석 대상 종목에 대한 정보를 한눈에 확인할 수 있는 웹 서비스 구현** | |
| **개요** | 1. 데이터 수집  ① 크롤링으로 유튜브 스크립트, 언론 보도 이슈 수집  - 유튜브의 경우 ( 구독자수 기준 ) : 주식 관련 유튜브 채널 3개  - 언론보도 자료 ( 유료부수 구독자 수 기준 ) : 매일경제, 아시아경제  - 매일 업로드 되는 자료에 대한 스크립트 추출 예정  ② FinanceDataReader 라이브러리로 주식가격 데이터 수집 ( 한국거래소 자료 )  - 매일 실시간 종가 데이터 수집  - 시가 총액 상위 종목 선정  2. 데이터 전처리  ① 유튜브 스크립트 & 언론 보도 이슈  - 형태소 분석 · 불용어 처리  - 휴장일 날짜 조정  ② 주식가격  - 종목마다의 가격 차이를 고려하여 MinMaxScaler 적용  3. 데이터 분석  ① 유튜브 스크립트 & 언론 보도 이슈  - 주가 등락 예측을 위한 감성 분석  ② 주식가격  - LSTM · ARIMA · FBProphet · 강화 학습으로 주식가격 등락 예측  ③ Label  - ①+② 분석을 종합하여 주식가격의 상승/보합/하락 예측  4. 데이터 시각화  - 일별 이슈 키워드를 나타낸 워드 클라우드로  - 긍정/부정 감성점수를 통한 뉴스 표출  - 주가의 시계열 그래프  5. 웹 페이지 구현  - 이슈를 한 눈에 볼 수 있도록 웹 페이지 구현  - 해당 종목의 주식 그래프(전날 종가) 및 주식가격 예측 결과 표출  - 유튜브 이슈 워드 클라우드  - 긍정/부정 뉴스에 대한 서비스 사용자의 판단 수렴 기능 | |
| **프로젝트 수행 방향**  **(주요 기능 설명)** | **사이언스** | **엔지니어링** |
| - 텍스트 데이터 EDA 및 텍스트 벡터화  - 감성 사전 활용 및 제작  - 텍스트 감성 분석  - LSTM · ARIMA · FBProphet ·   강화 학습 모델링  - 감성 분석 결과와 다른 모델 결과를 결합하여 최종 앙상블 모델링 | - 데이터 전처리  - 데이터 수집  - Tableau  - 웹 페이지 연동  - 백엔드 구현 |
| **프로젝트 수행 도구** | **사이언스** | **엔지니어링** |
| - Python  - TensorFlow  - Tableau | - Python  - JavaScript  - Django  - MySQL |

| **필수 기능** | **사이언스** | **엔지니어링** |  |
| --- | --- | --- | --- |
| * 미디어 키워드 빈도 분석 시각화 * 키워드 감성 분석 결과 시각화 * LSTM · ARIMA · FBProphet  · 강화 학습 예측 주가 그래프 * 최종 예측 결과 표출 | * 데이터 크롤링 * 시각화 결과 표출 * 대시보드 * 웹 페이지 구현 |  |
| **포함 기술** | **사이언스** | **엔지니어링** |  |
| * 딥러닝 (LSTM) * 머신러닝 (ARIMA, FBProphet) * 강화 학습 * 워드 클라우드 * 텍스트 마이닝 * 감성분석 | * 크롤링 데이터 DB 저장 * 비정형 데이터 처리 기술 * 정형 데이터 변환 * 데이터 파이프라인 구축 |  |