

주제: 뉴스 키워드 기반 주가 등락 여부 예측

문제 정의:

젊은 세대의 주식 투자에 대한 관심 증가

목표:

주식 투자에 대한 수익률 증가

실행 계획:

1. 뉴스 기사(또는 커뮤니티 등), 주가 정보 크롤링(네이버, 구글 등) 및 수집
 - Python 사용하여 비동기 방식으로 수행 후 HDFS에 저장 / Spark 활용하여 병렬 수행?
2. 텍스트 데이터 전처리(불용어 제거 등)
 - Spark 활용하여 병렬로 수행
3. 뉴스 기사 감성 분석(긍정/중립/부정)
4. 감성 분석 정보 및 기타 정보 등 활용하여 등락 예측
 - 실시간 스트리밍이 필요하다면 Kafka 활용

+ 문제점

난이도, 성능("주가"라는 것 자체가 예측하기 어려움)

++ 기존에 생각했던 내용은 현실적으로 어려울 것으로 생각되어, 감성 분석으로 변경