

FinTech-AI 해커톤 기획서

- 일반 개인 투자자를 위한 산업군별 기사 분류 및 부정 기사 알림 서비스 -

팀명: FINTACK

임태민

엄세웅

한동훈

1. 팀 소개

임태민 (조장): hugeman@khu.ac.kr

- 경희대학교 국제학과/컴퓨터공학과 4 학년 재학
- "의류 온라인 상품 후기 객관화를 위한 온라인 쇼핑몰 댓글 리뷰 및 분석", 한국정보과학회 2019 한국소프트웨어종합학술대회 VOL 46 NO. 02 PP. 1492 ~ 1494 2019.

엄세웅: exopas@khu.ac.kr

- 경희대학교 국제학과/컴퓨터공학과/ 소프트웨어융합학과 데이터사이언스 트랙 4 학년 재학
- Spirent Communication Silicon Valley Engineering Intern (2019.08~2020.02)
- 서울대학교 융합기술대학원 UX Lab 학부 연구생 (2020.07.01~)
- "의류 온라인 상품 후기 객관화를 위한 온라인 쇼핑몰 댓글 리뷰 및 분석", 한국정보과학회 2019 한국소프트웨어종합학술대회 VOL 46 NO. 02 PP. 1492 ~ 1494 2019.

한동훈: dhh96@khu.ac.kr

- 경희대학교 국제학과/컴퓨터공학과 4 학년 재학
- Best of the Best 8 기 보안제품개발트랙 수료
- 서울대학교 융합기술대학원 Applied Data Science Lab 학부 연구생 (2020.08.01~)
- "의류 온라인 상품 후기 객관화를 위한 온라인 쇼핑몰 댓글 리뷰 및 분석", 한국정보과학회 2019 한국소프트웨어종합학술대회 VOL 46 NO. 02 PP. 1492 ~ 1494 2019.

2. 서비스 기획 의도

2019년을 기점으로 대한민국 전체 인구 대비 주식투자자 비율이 10%를 넘겼다. 통계청과 행정안전부의 분석에 따르면 경제활동인구 2758만명 중에서 20%에 해당하는 55만명이 주식투자를 하고 있으며 이는 돈을 버는 사람 5명 중 1명은 주식 투자를 하고 있다는 것을 의미한다. 경제활동과 주식 투자를 병행하는 일반 개인 투자자들이 점점 늘어나면서 매일 급변하는 주식 시장에서 빠르게 정보를 전달해주고 분석해주는 서비스에 대한 수요 또한 증가하고 있다.

학생 혹은 직장인 신분으로 매일 새로 생겨나는 다양한 기사를 취합하고 분석하는 것은 많은 에너지를 소모하는 일이다. 하나의 주식에 관심이 있다면 다행이지만 바이오 혹은 전기 자동차와 같이 산업군에 대한 정보를 취합하고자 한다면 훨씬 더 많은 시간을 쏟아야 한다. 주식과 관련된 기사를 알려주는 주식 애플리케이션에서도 크게 뉴스별 기사, 실시간 검색이 가장 많이 된 기사, 조회 수가 가장 많은 기사를 알려줄 뿐, 산업군별 중요 기사를 보여주는 기능은 존재하지 않아서 결국 별도로 기사를 찾아보고 취합해야 한다.

이와 더불어 실시간으로 주식 트렌드를 따라가기 어렵다는 문제점도 있다. 주식 장이 활성화되는 시간은 오전 9시에서 오후 3시 30분까지로 일반 개인투자자들이 직장에서 일을 하는 시간과 겹친다. 그 때문에 자신이 가지고 있는 주식에 관한 부정적인 기사가 올라와도 즉각적으로 대처하기가 힘들다. 여러 서비스에서 자신이 가지고 있는 주식과 관련된 기사를 알려주는 기능은 제공하고 있지만, 부정적인 기사만을 특정해서 알려주는 기능은 가지고 있지 않다. 개인 투자자들이 다른 일을 하고 있더라도 부정적인 기사가 나왔을 때 이를 즉각적으로 알려주는 기능이 있다면 개인 투자자들이 투자하는데 도움을 받을 수 있을 것이다.

3. 서비스 목적

본 서비스는 경제활동과 주식 투자를 병행하는 일반 개인 투자자들을 주 타깃으로 한다. FINACK 팀은 1) 산업군별 기사 분류, 2) 지정한 주식에 대한 부정 기사 알림 제공 기능을 제공하며 이를 통해 개인 투자자들의 편의성을 증진하는 것을 목표로 한다.

먼저, 산업군별 기사 분류 서비스는 산업군과 관련된 기사와 정보 등을 한눈에 파악할 수 있도록 정보를 제공하고자 한다. 이전에는 본인이 관심 있는 산업군별 정보를 파악하기 위해서 종목별 기사를 검색하고 취합해야 했기 때문에 많은 시간을 소모해야 했다. 하지만 해당 기능을 사용하면 별도의 노력이나 작업 없이 퇴근 후, 혹은 출근 전에 간편하게 산업별로 분류된 기사를 확인할 수 있다. 산업군 별 기사 분류 서비스는 주식 투자를 병행하고 있는 바쁜 개인 투자자들에게 있어서, 주식과 관련된 정보를 취득하는데 소요되는 시간을 획기적으로 줄여줄 수 있을 것이다.

다음으로, 지정한 주식 종목에 대한 부정적인 기사가 발표되었을 때 이를 알려주는 팝업 알림 기능을 제시하고자 한다. 많은 개인 투자자들은 직장에서 근무하느라 자신이 투자하는 종목에 대한 소식을 놓치는 경우가 많다. 이로 인해 막대한 손실을 보는 경우가 많은데, 이런 상황에서 팝업 알림 서비스는 개인 투자자들이 손해를 보는 것을 방지하고 즉각적인 대응을 하는 데 도움을 줄 것이다. 나아가 근무 중 손쓸 수 없는 주식의 급변 가능성에 대한 심리적 불안감을 개선하며 개인 투자자들이 더 원활하게 투자를 하는 데 도움을 줄 것이다.

4. 차별화

현재 주식 관련 애플리케이션에서 주식 관련 기사를 제공하는 방식은 크게 세 가지가 있다. 먼저 모든 기사를 나열해서 보여주는 방식과 자신이 매수한 주식과 관련된 기사만 실시간으로 업데이트해주는 방식이 있다. 다음으로 실시간 검색어처럼, 최근 이슈화되고 있는 기사를 제공하는 방식이 있다. 예를 들어 키움증권에서 서비스하고 있는 영웅문 S 애플리케이션에서는 전체 기사를 보여주는 종합 뉴스, 이슈화되고 있는 기사를 보여주는 마켓 N 이슈, 종목별 뉴스를 알려주는 종목 뉴스 등이 있다.

FINTACK 팀은 위와 같이 단순히 정보를 나열해주는 정보 제공 방식과는 다른 방식으로 더욱 양질의 정보를 제공하고자 한다. 먼저 대표적인 AI 기술인 자연어 처리를 사용하여 특정 기간(하루, 일주일) 동안 화제가 된 뉴스들을 산업군별로 나눠서 제공하고자 한다. 하나의 기사에서 산업군 키워드를 찾아내고 이를 '에너지', '소재' 등의 25 가지 산업군 카테고리에 할당하여 산업군별로 이슈화 되는 기사를 요약해서 볼 수 있게 제공할 것이다.

또한 기사 정보를 취합하고 산업별로 분류해주는 것에서 더 나아가 경제 활동과 투자를 병행하는 개인 투자자들이 부정적인 이슈에 즉각적으로 대처할 수 있도록 하여 차별점을 두고자 한다. 사용자가 매수한 종목을 설정해 놓으면 해당 종목에 대한 부정적인 기사가 발표되었을 때 팝업을 통해 알려주는 서비스를 제공할 것이다. 자연어처리 기법과 크롤링을 활용하여 새로 발표되는 기사를 선별하고 자연어 처리 기법의 하나인 감정 분석을 통하여 기사의 긍정/부정 여부를 파악한다. 이때 기사가 부정적이라면 팝업이나 알림을 통해서 투자자가 즉각적으로 대처할 수 있도록 도와주는 서비스를 제공하고자 한다.

5. 데이터

1. 뉴스 데이터

네이버 금융 실시간 속보에 올라오는 뉴스 기사를 크롤링하여 기사의 헤드라인과 요약 정보를 수급할 예정이다.

(참고 링크:

https://finance.naver.com/news/news_list.nhn?mode=LSS2D§ion_id=101§ion_id2=258

2. 코스피 200종목 리스트 & 산업군 라벨

코스피 200종목 리스트는 네이버 금융에서 제공하는 코스피 200 종목 구성을 참고한다. (참고 링크: <https://finance.daum.net/domestic/kospi200>)

산업군 라벨은 에프앤가이드에서 제시하고 있는 방식을 따라 에너지, 소재, 자본재 등의 25가지 산업군 라벨을 사용한다. (참고 링크: <https://www.fnguide.com/>)

6. 알고리즘

1. 산업군 별 기사 분류: 다중 클래스 분류(Multiclass classification)

기사들을 25 가지 산업군으로 분류하기 위하여 다중분류 알고리즘(Multiclass classification)을 사용한다. 먼저 크롤링을 통해 얻은 뉴스 데이터에서 KR-WordRank 와 같은 비지도학습 방법의 키워드 추출 알고리즘을 사용하여 키워드를 추출하고 이를 다중 분류 알고리즘을 통해 분류하여 각 기사가 어느 산업군에 속하는지 파악한다.

2. 지정한 주식에 대한 부정 기사 알림 기능: 장단기 메모리(LSTM)를 사용한 감정분석(Sentiment analysis)

기사가 부정적인 내용을 포함하는지를 파악하기 위해 자연어 처리 기법의 하나인 장단기 메모리(LSTM)를 활용한 모델을 사용한다. 크롤링한 기사에서 KR-WordRank와 같은 비지도 학습 방법의 키워드 추출 알고리즘을 사용하여 등록한 주식 종목과 관련된 기사인지 먼저 파악한다. 이후 기사의 제목과 요약 내용을 LSTM을 활용한 모델을 사용하여 긍정/중립/부정 여부를 파악하고 부정으로 판별된 기사의 경우 팝업을 통해 사용자에게 정보를 전달한다.