AI 로보어드바이저를 통한 ETF 추천 서비스

이은주[1], 박설하[1], 이승준[2], 이예령[3], 문재현*
[1]명지대학교 산업경영공학과
[2]전남대학교 소프트웨어공학과
[3]인하대학교 수학과
*한국기술거래사회
eunju_2483@naver.com, tjfgk709@gmail.com, dltmdwns3462@naver.com, ery270@naver.com, smjhoon@gmail.com
*교신저자(Corresponding Author)

ETF Recommendation Service through AI RoboAdvisor

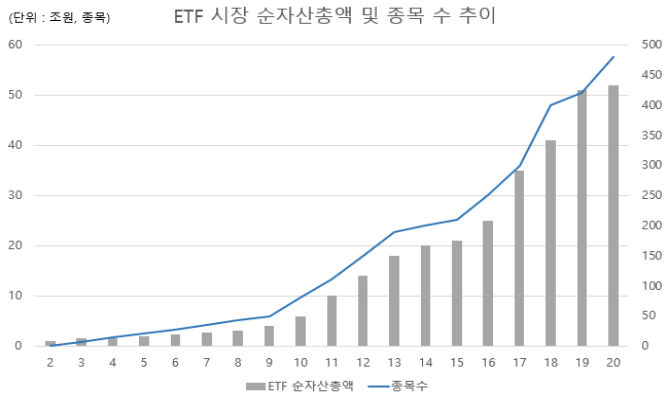
Eun-Ju Lee[1], Seol-Ha Park[1], Seung-Jun Lee[2], Ye-Ryung Lee[3], Jae-Hyun Moon*
[1]Dept. of Industrial Management Engineering, Myong-Ji University
[2]Dept. of Software Engineering, Chon-Nam National University
[3]Dept. of Mathematics, In-Ha University
*Korea Technology Transfer Agent Association

요 약

투자에 대한 관심 증가에 따라 적은 비용과 시간으로 객관적인 정보 제공의 필요성 증가와 함께 인공지능 기술을 활용한 로보어드바이저 서비스가 확대되었다. 또한, 최근 ETF 를 통한 안정적인 투자에 대한 선호도가 증가함에 따라 ETF 중심의 AI 로보어드바이저 추천 서비스가 필요할 것으로 보인다. 하지만, 기존의 투자 어플리케이션에서는 뉴스 기반의 감성적인 요인이 반영되지 않은 추천 방식으로 주가에 영향을 미치는 다양한 요인들을 고려하지 못하는 문제점이 있다. 이에 본 연구에서는 뉴스의 감성분석을 통한 감성지수를 기반으로 새로운 주가 예측 모델을 제안하고, 사용자의 투자 성향 분석을 통한 맞춤 추천 서비스를 통해 개인화된 ETF 서비스를 제공한다.

1. 서론

투자에 대한 관심 증가로 적은 비용과 시간으로 객관적인 투자 정보 제공의 필요성이 증가하였으며, 이에 인공지능 기술을 활용한 로보어드바이저 서비스가 도입 및 확대되고 있다. 국내 로보어드바이저 서비스 이용자 수는 꾸준히 증가해 2019년 9월 기준 10만 명을 돌파하였으며, 전 세계적으로는 2023년 로보어드바이저 시장 규모가 2조 5,523억 달러가 될 것으로 전망하고 있다.[1]



(그림 1) ETF 시장 순자산 총액 및 종목 수 추이

이와 더불어 최근 안정적인 투자에 대한 관심도 증가하면서 ETF의 순자산 총액이 50조 원이 넘어설 만큼 국내/해외 ETF 대한 관심과 투자가 증가하고 있으며, 다양한 테마형 ETF들이 출시되고 있다.

이에 따라 ETF 중심의 인공지능 로보어드바이저 추천 서비스가 필요할 것으로 보인다.

2. 문제 제기

지속적으로 증가하는 투자에 대한 수요를 충족시키기 위해 다양한 투자 정보를 제공하고, 실제 투자 혹은 모의투자가 가능한 어플리케이션이 많이 이용되고 있다. 하지만 기존의 투자 서비스의 경우, 해외 ETF에 대한 설명이 다소 부족하다. 또한, 투자 종목을 추천할 때에 뉴스 기반의 감성적 요인이 반영되지 않은 추천 방식이 사용되고 있다. 이러한 방식은 투자가 처음인 사용자의 경우 투자에 어려움을 겪는 상황이 발생할 수 있으며, 뉴스와 같이 주가에 영향을 미치는 다양한 요인들을 고려할 수 없다.[2]

뉴스는 현실 세계에 일어나는 각종 현상에 대한 설

명과 정보들을 포함하고 있기 때문에 주식가격 변동의 중요한 원인으로 작용한다.[3]

(그림 2) 주가에 영향을 미치는 근본 요인들

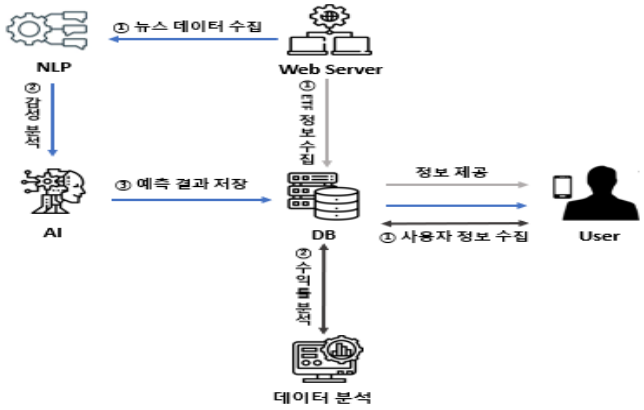
이에 본 논문은 주가 변동의 주요 요인인 뉴스의 감성분석을 통해 도출된 감성지수를 기반으로 새로운 주가 예측 모델을 수립하여 단순 예측이 아닌 긍정/부정 요소까지 고려한 예측 방법을 제안한다. 또한, 국내/해외 ETF 정보 및 관련 뉴스를 제공하고 사용자의 투자 성향 분석을 통한 맞춤 ETF 추천과 예상 수익률 제시를 기반으로 모의투자가 가능한 시스템을 구축하고자 한다.

3. 제안 시스템

3.1 시스템 구성

본 시스템은 서버와 스마트폰 어플리케이션, DB로 구성되어있다. 서버는 지속적으로 국내/해외 뉴스 및 ETF 주가 정보를 수집하여 DB에 저장한다. 저장된 뉴스를 'Okt 형태소 분석기'를 활용한 자연어 처리를 통해 감성분석을 수행하여 감성지수를 도출한다. 이후, 감성지수를 고려한 주가 예측 모델을 통해 ETF 주가를 예측하고 결과를 DB에 저장된다.

스마트폰 어플리케이션의 투자 성향 설문 및 로그인을 통해 사용자 정보를 수집하여 데이터베이스에 저장한다. 예측된 ETF 수익률과 투자 성향을 기반으로 사용자는 맞춤형 ETF를 추천 받고 모의투자 할 수 있으며, 해당 정보 역시 DB에 저장되어 사용자는 추천 ETF와 모의투자 현황을 지속적으로 관리할 수 있다.



(그림 3) 시스템 구성도

3.2 데이터 수집 및 정제

본 시스템에서 필요한 데이터는 투자성향 및 사용자 정보, 모의투자 현황 데이터, 국내/해외 ETF 정보, 뉴스, 감성지수 데이터이다. 사용자 정보와 투자 성향의 경우, 어플리케이션의 로그인 및 회원가입과 투자 성향 설문 기능을 통해 수집하였다. 투자 성향의 경우에는 사용자가 원할 때 새롭게 설문 가능하며 정보에 변동이 있을 경우, DB에 새로 저장된다.

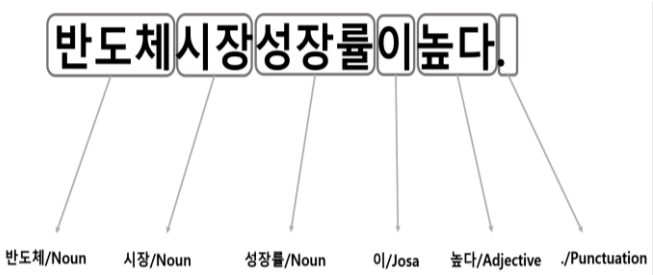
국내/해외 ETF 주가 데이터는 Open API를 활용하여 수집했으며, 뉴스 데이터는 웹 크롤링을 통해 수집했다. 두 데이터 모두 현재 날짜를 기준으로 3년 치 데이터를 수집했으며, 뉴스 데이터의 경우에는 수집 이후 자연어 처리를 활용한 정제 과정을 진행하여 이를 기반으로 감성 지수 데이터를 도출했다.

모의투자 현황 데이터의 경우, 사용자가 추천 받은 ETF 혹은 투자해보고 싶은 ETF를 직접 검색하여 매수/매도했을 때에 수집된다. 사용자의 ETF 매수/매도 시에 종목명, 가격, 수량 등이 DB에 저장되고 이전의 투자 내역이 갱신된다.

3.3 뉴스 기반 감성 분석

국내 뉴스가 국내 ETF 주가에 미치는 영향을 고려한 주가 예측 모델을 생성하기 위해 뉴스의 감성 분석을 통해 감성지수를 도출하였다. 웹 크롤링을 통해 수집된 뉴스 데이터를 활용하였으며, 뉴스의 내용뿐만 아니라 시간 정보까지 수집하여 변수로서 재가공하였다. 이후, 해당 뉴스의 긍정/부정을 판단하기 위해 뉴스 기사의 키워드를 기반으로 분석을 진행하였다.

키워드 분석은 'Okt 형태소 분석기'를 활용하여 텍스트를 형태소 단위로 분리하고 관사, 조사, 전치사 등의 불용어를 제거하여 극성을 판단할 수 있는 단어를 도출하였다. 이를 감성 사전과 비교하여 단어별 긍정/부정 극성을 판단하였다.[4]



(그림 4) 형태소 분석

구축되어 있는 기존의 감성 사전만 활용하기에는 주식 및 경제 관련 단어의 특성을 반영하기 어려울 것이라고 판단되었다. 또한 외래어 혹은 신조어로 이루어진 주식 용어가 다수 존재하며, 계속 생겨나기

때문에 이를 반영해야 한다고 생각되어 주식 용어와 관련된 주식 사전을 새로 구축하였다.

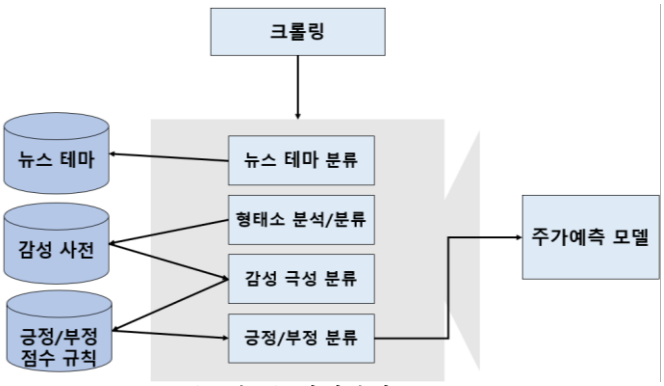
주식 사전은 단어별로 -2, -1, 0, +1, +2, 총 5가지 점수를 부여하는 방식을 사용하였다. 예를 들어, '급락'과 같이 분명하게 부정적임을 알리는 키워드의 경우 절댓값이 큰 음수인 -2를 부여하고 '성장'과 같이 뒤에 오는 문장(가능성이 떨어진다, 가능성이 높다)에 따라 판단이 상반되는 경우가 존재하는 키워드는 절댓값이 작은 양수 +1을 부여하는 방식으로 사전을 구축하였다.

직접 구축한 주식 사전과 기존 한국어 감성 사전을 활용하여 각 단어들의 긍정/부정 값을 뉴스의 긍정/부정 비율로 합산하여 (1)과 같은 방법으로 1개의 뉴스에 대한 긍정/부정 비율을 계산하였다.[4]

$$\text{NewsPNr} = \frac{\sum_{i=1}^{n} wordPN(i)}{n} \times 100\% \quad (1)$$

이와 같은 방법으로 3년 치의 모든 뉴스에 대한 일별 점수를 계산하여 일별 감성지수를 도출하였다. 도출된 감성지수 중 결측된 데이터(Missing Value)의 경우, 결측된 날짜의 하루 전날의 감성지수 값으로 대체하는 방식을 사용하였다.

직접 구축한 주식 사전과 기존의 감성 사전을 함께 사용함으로써 주식 및 경제 관련 단어들이 많은 ETF 관련 뉴스의 특성이 잘 반영되도록 보완하였다. 이를 통해 더 정확한 감성지수를 도출하였다.



(그림 5) 감성분석 흐름도

3.4 시계열 예측 인공지능 알고리즘 비교

시계열 기반의 ETF 주가를 예측하기 위한 인공지능 알고리즘으로 LSTM 모델과 CNN 모델을 비교하였다.

LSTM 알고리즘은 데이터에 존재하는 장기간의 연관관계를 잘 파악해내며, 관측치 간의 순서를 명시적으로 모형에 반영하여 입력과 출력 간의 함수를 학습한다는 장점이 있지만, CNN 알고리즘에서는 이러한 기능이 존재하지 않는다. 반면, CNN 알고리즘은 모형 전체가 원천 데이터로부터 예측에 가장 적합한 속성들을 자동적으로 추출해내는 장점이 있다.

두 알고리즘의 특성이 매우 상이하여 본 논문에서 제안하는 감성지수를 고려한 주가 예측 모델을 LSTM, CNN 알고리즘으로 각각 수립하여 평균 제곱 오차인 MSE를 기반으로 성능을 비교하였다. 이에 LSTM 알고리즘은 0.002, CNN 알고리즘은 0.01으로 LSTM 알고리즘 기반의 모델이 더욱 우수한 예측 성능을 보였다.

또한, 1일 단위의 감성 분석 평균값이 시계열 데이터의 형태로 나타나기 때문에 LSTM 알고리즘이 이를 다루는데 특화 되어있으므로 이와 같은 특성과 MSE의 결과에 따라 LSTM 알고리즘을 선택하였다.[5]

3.5 감성지수를 고려한 주가 예측 모델 수립

국내 ETF 주가 예측 모델의 생성을 위해 뉴스를 기반으로 도출된 감성지수를 ETF 주가 데이터(종가, 고가, 저가, 거래대금)에 결합하였다. 이후, 결합된 데이터를 모델에 학습시키기 위하여 데이터 정규화 과정을 진행하였다.

정규화한 데이터는 7:3의 비율로 Train data와 Test data로 분리하고, TensorFlow Dataset을 활용해 지난 30일을 학습시켜 이후 5일 동안의 ETF 종가를 예측하도록 지정한 시퀀스 데이터 셋을 구성했다.

LSTM 모델을 기반으로 정규화한 3년 치 데이터를 학습시켜 뉴스 기반의 감성적 요인을 고려한 새로운 ETF 주가 예측 모델을 생성하였다.

해외 ETF 주가 예측 모델의 경우는 ETF 주가 데이터만을 활용하여 국내 ETF 추천 모델과 동일한 방법으로 생성하였다.

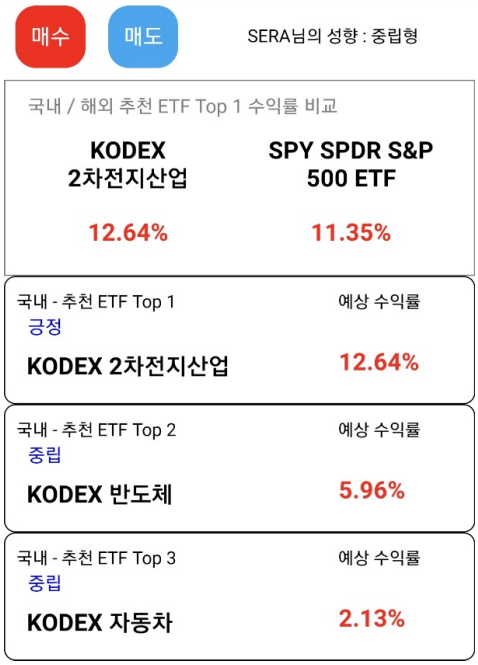
국내 ETF 주가 예측 모델 생성 시, 감성지수라는 새로운 변수를 추가하여 생성한 LSTM 기반의 새로운 예측 모델을 통해 MSE가 0.0047에서 0.002로 기존 모델보다 성능이 더 향상되었다.

3.6 사용자 맞춤 ETF 추천 및 모의투자

맞춤 ETF 추천을 위한 투자 성향 설문을 통하여 수집한 사용자의 투자 성향(공격형, 중립형, 안정형)과 일치하는 ETF들의 주가를 앞서 생성한 모델을 기반으로 예측한다. 사용자에게 ETF를 추천하는 날(모델 학습 날)을 기준으로 이후 5일의 ETF 종가를 예측하고, 예상 수익률을 계산한다.

단순히 예상 수익률을 기준으로 하여 ETF를 추천할 경우, 해당 종목이 상승 구간인지 하강 구간인지를 고려할 수 없다고 판단하여 이동평균을 이용해 해당 문제점을 보완하였다. 이를 기반으로 5일 뒤 예측 수익률 기준 상위 3개의 ETF를 사용자에게 추천할 수 있도록 하였다. 맞춤 ETF 추천 시에는 종목명 뿐만 아니라 예상 수익률, 감성분석 결과(긍정, 중립, 부

정)에 대한 정보를 함께 제공하여 모의투자를 위한 사용자의 판단에 더 많은 도움을 줄 수 있도록 한다.



(그림 6) ETF 추천 및 모의투자 화면

4. 결론

본 논문에서는 뉴스의 영향을 고려한 주가 예측 모델을 제안하고, 해당 모델과 투자 성향에 따른 사용자 맞춤 ETF 추천 시스템을 구현하였다.

본 논문의 시스템에서는 뉴스 기반의 감성분석 시, 기존의 감성 사전과 함께 주식 사전을 직접 구축하여 사용함으로써 주식 및 경제 관련 단어들이 많은 ETF 관련 뉴스의 특성 뿐만 아니라 외래어, 신조어로 이루어진 주식 용어들의 특성까지 잘 반영되도록 보완하였다.

또한, 감성지수를 더한 LSTM 기반의 새로운 주가 예측 모델을 활용함으로써 주가에 영향을 미치는 감성적 요인까지 고려할 수 있도록 하여 사용자에게 더 정확한 예측 수익률 정보를 제공할 수 있다.

이를 통해 사용자에게 예상 수익률, 감성분석 결과(긍정, 중립, 부정) 등 기존의 서비스보다 더 다양하고 정확한 정보를 제공한다. 더불어 국내/해외 ETF에 대한 정보와 모의투자 기능을 통해 초보 투자자의 진입 장벽을 낮추고 투자 안목을 기르는 데 도움이 될 것이라고 판단된다. 또한, 사전에 사용자의 투자 성향을 파악하여 사용자의 성향과 상황을 고려한 개인화된 서비스를 제공할 수 있다.

본 연구에서는 국내 뉴스에 대한 감성 분석을 통해 국내 ETF에 대한 새로운 주가 예측 모델을 한정적으로 수립했지만, 추후 해외 뉴스의 감성 분석까지 진행해 이를 해외 ETF 주가 예측 모델에 반영한다면 시스템을 더욱 발전시킬 수 있을 것이라고 판단된다.

또한, ETF 추천 서비스로 국한 되어있지만, 감성지수를 고려한 주가 예측 모델과 사용자 맞춤 추천 시스템을 활용하여 다양한 서비스를 개발할 수 있을 것이다.

참고문헌

[1]정한민, 황미녕. “인공지능 기반 로보어드바이저 운용 및 기술 동향”.ITFIND.2020

[2]김유신, 김남규, 정승렬. “뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형”.한국지능정보시스템학회.제 18권.제 2호 p143~156.2012

[3]차은정. “뉴스 감성분석과 SVM을 이용한 주가지수 예측”. 부산대학교.2016

[4] 김유신, 김남규, 정승렬. “뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형”.한국지능정보시스템학회.제 18권.제 2호 p143~156.2012

[5]강민규, 김보선, 신무곤, 백의준, 김명섭. “LSTM 기반 감성분석을 이용한 비트코인 가격 등락 예측.한국통신학회. p561~562.2020

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.