

인공지능 로보어드바이저를 통한 재테크 예측 시스템 구축에 관한 연구

김민기*, 김연수*, 양정우**, 조선근*, 문재현***⁺

*울산과학기술원 경영학부

**울산과학기술원 자연과학부

***한국기술거래사회, ⁺교신저자(Corresponding author)

mingi3314@unist.ac.kr, yeon17@unist.ac.kr, wjddn955@unist.ac.kr, sk37025@unist.ac.kr, smjhoon@gmail.com

A Study on Building a Financial Prediction System with Artificial Intelligence Robo-Advisor

Minki Kim*, Yeonsu Kim*, Jeong-Woo Yang**, Sunkeun Jo*, Jaehyun Moon***

*School of Business Administration, UNIST

**School of Natural Science, UNIST

*** Korea Technology Transfer Agents Association

요약

국민연금이 2056년 고갈될 수 있다는 전망이 나오면서 연금소득에 대한 국민들의 불안감이 커졌다. 노후를 위해 미리 대비해야한다는 인식이 커지며 자동으로 투자해주는 '로보어드바이저'에 대한 사회적 관심이 함께 높아졌다. 본 연구에서는 기존 시중 은행들의 펀드 기반 로보어드바이저가 아닌 기업 재무 정보, 수정 종가 데이터를 이용한 직접 투자를 고안하였다. LGBM 알고리즘으로 포트폴리오를 구현해본 결과 실제 퀀트 투자에서 사용되는 지표들이 주식의 변화를 예측하는데 효과가 있음을 확인할 수 있었다.

1. 서론

최근 국민연금이 2056년 고갈될 것이라는 전망이 나오며 현재 세대가 미래에 연금을 보장받기 어려울 수 있다는 불안감이 퍼졌다. 국민연금공단에서는 2018년 16개 지역에서 국민 토론회를 개최하여 '국민연금종합운영계획'을 수립하기 위해 노력하였다. 개인이 안정된 노후를 위해 준비해야 한다는 인식이 증가하면서 개인연금, 퇴직연금, 국민연금으로 구성된 선진국형 3총 노후소득 보장체계에 대한 관심이 커졌다.

본 연구에서는 개인의 노후를 보장하기 위한 방안으로 로보어드바이저를 이용한 개인연금 운용을 제시하고자 한다. 그림 1에서와 같이 국내 로보어드바이저 시장 규모는 2016년 2,017억 원에서 2021년 1조 9,021억 원으로 5년내 약 9.4배의 성장률이 예상되고 있다. 로보어드바이저는 딥러닝을 기반으로 자동으로 펀드를 운용해주는 시스템으로 데이터에 근거한 객관적인 투자를 가능하게 한다. 또한 수수료 부담을 낮추고 높은 수익률을 보장한다.

2. 기존 로보어드바이저의 문제점

기존 시중 은행들의 로보어드바이저는 기존의 펀드를 섞어 추천하기 때문에 펀드 수수료의 부담을 갖게 된다. 또한 충분한 고객 분석이 없어 5가지 내외의 설문에 의존하기 때문에 개인화된 포트폴리오 구성은 어렵다. '평생직장'의 개념이 사라지고 기술의 발전으로 급격히 자동화가 이루어지는 시점에서 개인의 소득 상황 또한 변화할 수 있는 위험성을



[그림 1] 국내 로보어드바이저 시장 규모 전망 [1]

갖는다. 이러한 상황에서 고객과 밀접하게 접촉하고 변동사항을 즉시 반영할 수 있는 개인연금 로보어드 바이저의 중요성이 커지고 있다. [2]

3. 해결책 제안

본 연구에서는 30년 이상의 장기간 투자를 요구하는 개인연금을 로보어드바이저가 자동으로 관리해주는 시스템을 구축하고자 하였다. 또한 기존 방식처럼 펀드를 섞는 방식이 아닌, 실제 회사들의 재무제표를 근거로 각 회사들의 주식을 직접 섞어 포트폴리오를 구축함으로써 고객들의 수수료 부담을 줄인다. 또한 고객들의 투자기간, 최대 수용 가능 위험과 같은 투자성향을 바탕으로 개개인에게 최적화된 포트폴리오를 제공한다.

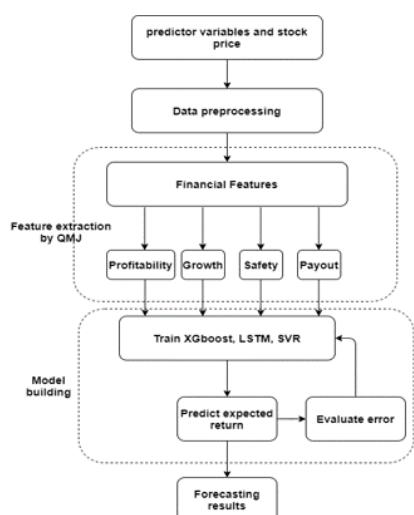
4. 인공지능기반의 자산관리 서비스

1) 시스템 개요

해당 시스템은 인공지능을 통해 예측된 기대수익률을 기반으로 최적화된 포트폴리오를 사용자에게 제공한다. 더불어 사용자의 성향에 따라 차별화된 투자 전략을 제공할 수 있도록 위험 선호도를 고려한 맞춤형 투자 제안을 추천한다. 또한 지속적인 성과 추적 및 피드백을 통해 장기적인 자산관리 서비스를 제공한다. 예측 알고리즘으로는 LGBM을 사용하였다.

2) 예측 인공지능 구성도

해당 시스템에 사용된 인공지능은 기업의 재무정보 및 주가 데이터를 이용해 기대수익률을 예측하도록 활용된다. [그림 2]은 예측 알고리즘의 흐름도이며 수집된 데이터는 전처리 과정을 거친 후 특징 추출을 통해 기업의 수익성, 성장성, 안정성, 지불금 요소를 생성한다. 생성된 요소들은 인공지능 모델의 입력 값으로 사용되며 모델은 성능이 일정 기준이 될 때까지 학습과정을 거친 후 기대수익률을 예측함으로써 마무리된다.



[그림 2] 알고리즘 흐름도

3) 데이터 수집 및 가공

알고리즘에서 필요한 데이터는 기업 재무 정보, 수정 종가 데이터이다. 우선 기업 재무 정보로는 각 분기별로 공시된 기업별 IFRS 연결 재무제표 데이터를 KOCOinfo에서 제공한 데이터베이스를 이용해 수집했다. 수정 종가 데이터는 DART에서 제공한 openAPI를 이용한 웹크롤러로 수집했다. 수집된 데이터는 전처리 과정을 거친 이후 GPOA, ΔGMAR, Ohlson's O-Score, EISS 등 기업의 수익성, 성장성, 안정성, 지불금을 나타낼 수 있는 20여 가지의 지표를 생성하는데 사용된다. 분기별로 정렬된 기업의 지표들은 t+1 시점의 수정 종가 데이터와 결합해 최종적으로 모델 훈련을 위한 가공을 마친다. [3]

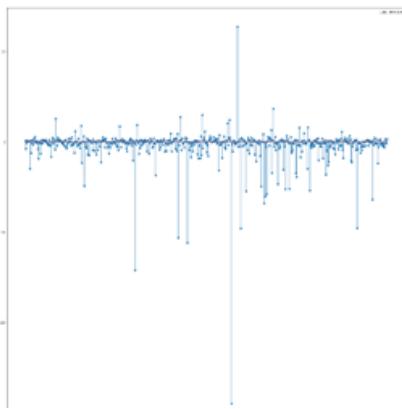
4) 알고리즘 설명

본 연구에서 사용한 알고리즘은 xgboost, Light Gradient Boosting Model(LGBM), Support Vector Regressor(SVR), LSTM이다. 이를 위해서 위의 데이터 수집 및 가공에서 모델을 위한 가공을 마쳤다. 가지고 있는 데이터의 특성상 결측 값이 존재할 수밖에 없는데, 결측 값이 존재하는 경우 트리(tree) 기반의 모델이 아닌 SVR의 경우 결측치를 채울 것인지 등의 판단이 필요하였다. SVR의 경우에는 알고리즘을 사용하기에 적합하지 않다고 판단하여 학습에서 제외하였다. [4]

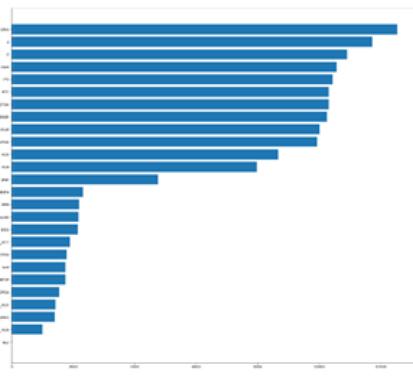
이와 다르게 트리 기반의 모델의 경우에는 결측 값이 있으면 해당 샘플들만 모아서 따로 진행하게 된다. 따라서 데이터 특성상 트리 기반 앙상블 모델인 LGBM, Xgboost 모델을 고려하였다. 두 모델은 boosting 모델이라는 공통점을 갖는다. 두 가지의 경우 모두 좋은 성능을 보여주지만 xgboost의 경우 하이퍼파라미터도 많으며 메모리를 많이 요구하는 단점을 갖고 있다. 최종적으로 비슷한 성능을 가지면서 메모리를 덜 요구하는 LGBM을 사용하였다.

LGBM의 경우 기존의 트리 BOOSTING 모델과 다르게 리프 중심 트리의 분할을 사용하여 리프 노드를 지속적으로 분할해 나아가기 때문에 동일한 리프(예측 값)을 나타낼 때 훨씬 더 손실을 줄일 수 있다는 장점이 존재한다. 추후 LSTM을 이용하여 예측 시점 이전까지의 데이터를 학습하여 3개월 이후의 데이터를 예측하고, 해당 값을 다시 트리 모델에서 다음 시점의 값을 예측하는데 사용한다면 트리 모델이 그 전의 변동성까지 충분히 학습하여 다음 값을 예측할 수 있을 것이라고 기대하며 알고리즘을 구축하였다. [5]

5) 알고리즘 성능



[그림 3] 알고리즘 결과 1



[그림 4] 알고리즘 결과 2

시계열 특성을 factorization 하여 LGBM 모델에 값을 넣고 평가 방법을 MAPE(평균 절대 백분율 편차)로 하여 결과를 분석해봤을 때 545 개의 무작위 시점의 무작위 회사에 대해 평균 76 퍼센트의 차이를 보여주었다. 그림 3,4 를 통해 결과를 시각화해 본 결과 평균 76 퍼센트의 차이를 보인 이유에 대해 간략하게 확인할 수 있었다. 일부의 값들이 학습이 진행되지 않아 실제 값과의 차이가 커기 때문에 이런 경우 MAPE 값을 높였다고 생각할 수 있다. 이는 추후 LSTM을 통해 해당 회사의 해당 시점 전의 트렌드를 반영한다면 충분히 개선의 여지가 있다고 판단된다.

[6]

LGBM 을 통해서 확인할 수 있었던 것은 실제 퀀트 투자에서 사용하는 지표인 지불금을 나타내는 지표 중 하나인 diss(부채의 변화 퍼센트) 와 다른 지표들을 합산하여 만들어낸 지표인 z-score 등 실제 퀀트 투자에서 사용되는 지표들이 주식의 변화를 예측하는데 효과가 있었다는 것이다. alpha 와 beta 같이 시간의 변동성을 나타내는 지표들의 경우 예측하는데 큰 도움이 없었다는 것을 확인할 수 있었다.

6) 예상 시나리오 설계

로보어드바이저 이용자의 투자 성향을 안정형, 위험 중립형, 적극투자형, 공격투자형의 4 가지로 나누어 개인에게 적합한 포트폴리오를 제시한다. 연금의 목적을 고려하였을 때, 연금은 수익률에 있어 기준의 투자와 다른 특이점을 가진다. 개인연금 투자는 원금이 보장되지 않는 상황에 대해 큰 거부감을 갖는다. 그러나 목표 수익률 이상을 달성했을 경우에도 투자자에게 디메리트(demerit) 되지 않는다. 따라서 가장 수익률이 높은 공격투자형이라 하더라도 보수적이고 안전한 운용이 중요시된다.

로보어드바이저를 통한 개인화된 연금 운용을 위해 웹페이지로 설문조사를 실시한다. 이용자의 나이, 소득 수준, 희망 생활 수준, 선호 투자 종목, 금융 지식 등의 정보들을 제공받는다. 이 정보들은 이용자의 투자 성향을 산출하기 위해 이용된다. 항목당 특정한 점수를 매겨 높은 점수 순으로 공격투자형, 적극투자형, 위험중립형, 안정형으로 구별된다.

[그림 5] 고객 투자 성향 설문조사 웹페이지

포트폴리오를 구성하기 위해서 Markowitz 의 포트폴리오 이론을 따른다. 가장 효율적인 포트폴리오 하나와 무위험자산인 채권(T-bill)을 결합하여 이용자의 위험 투자 성향에 따라 수익률과 분산을 조정한다. 이는 포트폴리오를 다수 구성하지 않아도 되는 장점을 갖는다.[7]

5. 결론

본 연구에서는 효율적인 개인연금 운용을 위한 로보어드바이저와 알고리즘 구축에 대해 알아보았다. 펀드 기반의 로보어드바이저가 아닌 재무제표와 투자 지표들을 이용한 주식 비율 배분 투자를 통해 수수료 부담을 줄일 수 있다. 이를 통해 사람들에게 더욱 유용한 로보어드바이저를 제공할 수 있게 되며 국민의 연금소득 보장에 기여할 수 있을 것이다.

참고문헌

- [1] 이코노마드(2017). 로보 어드바이저 완벽 개념정리, 뱅크샐러드 매거진.
- [2] 김선웅(2019). 지능형 전망모형을 결합한 로보어드 바이저 알고리즘. 한국지능정보시스템학회, 25(2), 39-55.
- [3] 이우식(2017). 딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측. 한국데이터정보과학회, 28(2), 287-295.
- [4] 송성환 외(2017). 머신 러닝 기반 소셜 빅데이터 분석을 이용한 금융자산 트레이딩 모델의 성능 향상에 관한 연구. Entrue Journal of Information Technology, 16(1), 51-62.
- [5] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780. 1997.
- [6] Tealab, A. (2018). Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*, 3(2), 334-340. 2018
- [7] Markowitz, H. (1952). Portfolio selection. *The Journal of FINANACE*, 7(1), 77-91

본 논문은 과학기술정보통신부 정보통신창의인재 양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.