

머신 러닝 모델을 활용한 주가 예측 및 트레이딩 전략의 성과 분석 연구

김진주, 김재윤*

순천향대학교

wlswn@sch.ac.kr, *kimym38@sch.ac.kr

A Study on the Performance Analysis of Stock Price Prediction and Trading Strategies Using Machine Learning Models

Kim Jinju, Kim Jaeyun*

Soonchunhyang Univ.

요약

본 연구는 코스피 시가총액 상위 40 개 종목의 데이터를 활용하여 주가 예측 및 투자 수익률 극대화를 위한 다양한 머신 러닝 모델과 트레이딩 전략을 평가하였다. 2010년부터 2024년까지의 데이터를 바탕으로 Random Forest, Decision Tree, XGBoost, LightGBM 모델을 학습하고 트레이딩 결과를 비교 분석하였다. 각 모델의 예측 성능과 다양한 트레이딩 전략의 효율성을 평가한 결과 LightGBM 모델과 두 번째 연속 신호에 따른 매매 전략이 가장 우수한 성과를 보였다. 본 연구는 주가 예측의 실용성을 높이고 투자 결정을 돋기 위한 최적의 모델과 전략을 제시하는 데 기여하며 연구의 결과는 향후 다양한 시장과 경제 환경에서 적용될 수 있는 보다 일반화된 예측 모델 개발의 필요성을 강조한다.

I. 서론

최근 몇 년간 주식 시장에 대한 관심이 증대되면서, 주식 시장 연구와 투자자의 수가 크게 증가하였다[1]. 하지만, 기존의 주가 예측 연구는 주로 모델 아키텍처와 예측 성능의 향상에 초점을 맞추고 있다[2]. 특히, 주식 예측 연구에서 중요한 요소는 높은 수익을 창출하는 것이다. 따라서 본 논문에서는 예측 결과를 활용한 전략 개발에 초점을 두었다. 주식 시장의 특정 패턴을 탐지하여 주가의 상승과 하락을 예측하고 다양한 분류 모델과 전략을 통해 투자 수익률을 극대화할 수 있는 방안에 대해 탐색했다. 최근 연구들에서 심층 학습 모델을 활용한 시계열 예측이 주식 시장의 복잡한 변동성을 효과적으로 포착할 수 있음이 입증되었다[3]. 이러한 접근은 투자 전략의 정밀성을 높이는 데 기여할 수 있다. 이를 고려하여 코스피 시가총액 상위 40 개의 종목 데이터를 사용하여 시장의 현재 상태와 향후 방향성을 예측하기 위해 다양한 기술적 지표를 계산하고 이를 모델의 입력 변수로 사용하였다. 기존 연구에서 Random Forest, XGBoost, LightGBM과 같은 머신 러닝 모델들은 주식 시장 예측에서 높은 성과를 보여왔으며, 이러한 모델들은 Profit Factor와 Payoff Ratio와 같은 성과 지표를 통해 평가된 바 있다[4]. 본 연구에서도 이러한 지표들을 활용하여 모델의 성능을 하여 수익률 극대화에 도움을 주며 합리적인 투자 결정을 할 수 있는 최적의 전략과 모델을 식별한다.

II. 본론

II.I 데이터 수집 및 전처리

본 연구에서는 yfinance 라이브러리를 사용하여 2010년부터 2024년까지 코스피 시가총액 상위 40 개 주요 종목의 데이터를 수집하였다. 수집된 데이터에는 종가, 고가, 저가, 시가, 거래량, 수정 종가의 주식 시장 지표가 포함되어 있으며 이를 기반으로 주식 및 금융 시장에서 가격과 거래량의 동향을 분석하였다. 추가적으로 주식 시장의 변동성을 반영하기 위해 다양한 기술적 지표들이 계산되었으며, 이러한 지표들은 모델의 입력 변수로 활용하였다. 데이터의 일관성을 유지하고 모델의 성능을 최적화하기 위해 훈련 데이터는 2010년 1월 1일부터 2020년 3월 16일까지의 데이터를 포함하며, 전체 데이터의 70%에 해당한다. 또한 목표 변수는 현재 시점보다 다음 날 주가의 상승(하강) 여부를 나타내도록 설정되었다.

II.II 학습 전략 및 평가 지표

본 연구에서는 다양한 머신 러닝 모델을 활용하여 주가 예측 모델을 구축하였다. Random Forest, Decision Tree, XGBoost, LightGBM 모델을 사용하였다. 모델 훈련 후, 테스트 데이터를 사용하여 주가 예측을 수행하였으며 목표 변수의 예측 결과를 바탕으로 6 개의 트레이딩 전략을 적용하였다. 이때 목표 변수는 현재 수정 종가보다 다음 날 수정 종가가 상승할 것이라고 예측하면 1, 하락할 것이라고 예측하면 0 으로 설정하였다. 또한 각 모델별로 Profit Factor(이익 계수), Payoff Ratio(지급 비율), Profit Average(수익 평균), Loss Average(손실 평균) 등의 성과 지표를 계산하여 모델의 성능을 평가하였다. 사용된 전략은 아래 표 2와 같다.

표 1. 트레이딩 전략

번호	전략 명칭
1	첫 번째 상승 또는 하락 신호가 발생했을 때 매수 또는 매도
2	두 번째 연속 상승 또는 하락 신호가 발생했을 때 매수 또는 매도
3	세 번째 연속 상승 또는 하락 신호가 발생했을 때 매수 또는 매도
4	세 번째 연속 상승 신호가 발생했을 때 매수, 두 번째 연속 하락신호가 발생했을 때 매도
5	네 번째 연속 상승신호가 발생했을 때 매수, 세 번째 연속 하락 신호가 발생했을 때 매도
6	Buy and Hold 전략, 예측 없이 매수 후 시장 변동에 따라 주식을 보유하여 유지

II. III 실험 결과

본 연구에서는 다양한 머신 러닝 모델과 트레이딩 전략을 활용하여 주가 예측과 투자 수익률을 평가하였다. 특히 전략 2(두 번째 연속 신호에 매매)에서 가장 우수한 성과가 나타났으며 이 전략의 결과를 중심으로 실험 결과를 분석하였다. 아래의 그림 1 박스 플롯은 각 모델별 Payoff Ratio 와 Profit Factor 의 분포를 시각적으로 보여주고 있다. 그래프에서 확인할 수 있듯이, LightGBM 모델이 다른 모델에 비해 더 높은 중앙값과 좁은 분포를 보여주면서 안정적이고 높은 성과를 나타낸다.

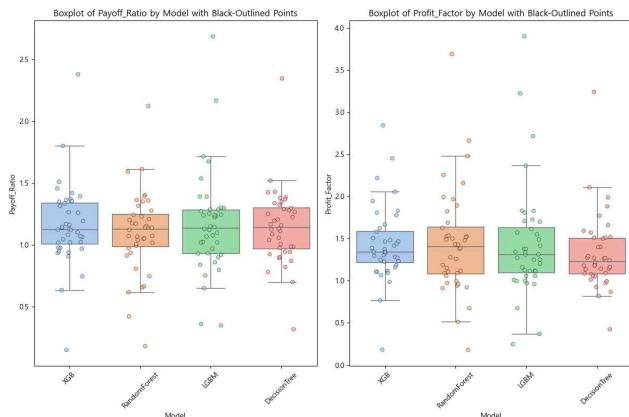


그림 1. 모델 별 Payoff Ratio 와 Profit Factor 의 분포

Decision Tree, XGBoost, Random Forest 모델 또한 유사한 성과를 보였으나 LightGBM 모델이 가장 높은 이익 계수와 지금 비율을 기록하였다. 이는 LightGBM 모델이 더 나은 예측 성능과 안정적인 수익 창출을 가능하게 했음을 의미하며 아래 표를 통한 이 결과들은 LightGBM 모델이 전략 2 를 기반으로 한 트레이딩에서 가장 효율적임을 보여준다.

표 1. 트레이딩 시뮬레이션 평균(표준편차) 결과값

Model	Profit Factor	Payoff Ratio	Profit Average	Loss Average
Decision Tree	1.33 (0.45)	1.14 (0.31)	4313.15 (5110.92)	-3807.66 (4347.47)
LGBM	1.44 (0.67)	1.16 (0.41)	4123.27 (4488.33)	-3676.88 (4077.66)

Random Forest	1.45 (0.62)	1.10 (0.33)	4049.9 (4250.97)	-3750.36 (3891.39)
XGB	1.43 (0.45)	1.16 (0.33)	4075.06 (4375.82)	-3703.96 (4060.64)

III. 결론

본 연구는 코스피 시가총액 상위 40 종목의 데이터를 활용하여 다양한 머신 러닝 모델과 트레이딩 전략의 성과를 비교 분석하였다. 실험 결과 LightGBM 모델과 두 번째 연속 신호에 따른 매매 전략이 가장 우수한 성과를 나타냈으며 이 모델은 다른 모델들에 비해 높은 Payoff Ratio 와 Profit Factor 를 기록하여 예측성능 뿐만 아니라 실제 투자 수익률에도 긍정적인 영향을 미쳤음을 확인할 수 있었다. 이러한 결과는 머신 러닝 모델이 금융 시장에서의 복잡한 패턴을 효과적으로 학습하고 예측하는 데 있어 높은 잠재력을 가지고 있음을 나타낸다. 그러나 본 연구는 코스피 시가총액 상위 40 종목과 특정 기간(2010-2024 년)에 국한된 데이터로 진행되었기 때문에 이 결과가 다른 시장이나 경제 환경에서도 동일하게 적용될 수 있을지는 불확실하다. 따라서 향후 연구에서는 다양한 시장과 경제 조건을 포함하여 더 긴 시간 범위를 다루는 등의 보다 종합적인 데이터를 통해 모델과 전략의 일반화 가능성 평가할 필요가 있다. 이를 통해 시장 특수성과 경제적 변동성을 고려한 더 신뢰할 수 있는 예측 모델과 트레이딩 전략을 개발할 수 있을 것으로 기대되며 이러한 추가 연구는 다양한 시장 조건에서 투자자들에게 실용적이고 적용 가능한 인사이트를 제공하는 데 기여할 것이다.

ACKNOWLEDGMENT

본 연구는 2024년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음. (2021-0-01399) 또한, 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1092808).

참 고 문 헌

- [1] Parveen, S., Satti, Z. W., Subhan, Q. A., & Jamil, S. (2020). Exploring market overreaction, investors' sentiments and investment decisions in an emerging stock market. *Borsa Istanbul Review*.
- [2] TJ Strader, JJ Rozicki, TH Root (2020). *Journal of Information Technology and Management*.
- [3] G. A. B. Abucay, K. C. C. Almonia, R. D. S. Buray, "Stock Price Predictor: Implementing Stocks Predictive Model Using Deep Learning," *International Journal of Computer Science and Research*, 2024.
- [4] J. Smith, A. Johnson, "A Comparative Study of Machine Learning Algorithms for Stock Market Prediction," *Journal of Financial Data Science*, 2022.