

딥러닝 기반 급등주 예측과 포트폴리오 수익률에 관한 연구

윤효경, 김정무*
영남대학교, *영남대학교

gyrud328@yu.ac.kr, *jungmu@yu.ac.kr

Deep Learning-Based Prediction of Price Jump and Portfolio Returns

Yun HyoGyeong, Kim Jungmu*
Yeungnam Univ., * Yeungnam Univ.

요 약

본 연구는 딥러닝 기법을 활용하여 국내 주식시장에서 급등주(limit-up stocks) 포트폴리오의 수익률을 분석하고, 이를 통해 효율적 시장 가설(Efficient Market Hypothesis, EMH)의 타당성을 재검토하고자 한다. 급등주는 단기간에 비정상적인 수익을 기록하는 종목으로, 투자자에게 높은 기회와 동시에 큰 위험을 제공한다. 기존 연구들은 급등주의 특성을 기술적 분석이나 전통적 계량 모형으로 설명하는 데 그쳤으나, 본 연구는 인공지능 기반 분석을 적용하여 예측 가능성을 실증적으로 검증하였다. 분석 결과, 급등주 예측은 완벽하지 않았음에도 불구하고, 예측 성공 종목(True Positive)이 제공하는 높은 수익률이 예측 실패 종목(False Positive)의 손실을 상쇄하며 전체 포트폴리오의 성과를 긍정적으로 이끌었다. 이는 개별 종목 차원의 정확도 한계를 넘어, 포트폴리오 차원에서는 딥러닝 기반 전략이 일정 수준 이상의 초과수익을 창출할 수 있음을 보여준다. 본 연구는 학문적으로 효율적 시장 가설의 약형(weak-form)에 대한 반론을 제기할 수 있는 근거를 제공하며, 실무적으로는 급등주 중심의 투자 전략이 가지는 가능성과 위험을 동시에 제시한다. 다만 분석 기간이 짧고 거래비용·유동성 제약을 반영하지 못한 한계가 존재한다. 향후 연구에서는 장기간 데이터를 활용하고 다양한 시장 상황을 고려하여 본 연구 결과의 일반화 가능성을 검증하는 한편, 보다 정교한 딥러닝 기법을 적용하여 투자 전략의 실질적 활용성을 심층적으로 탐구할 필요가 있다.

I. 서론

본 연구는 딥러닝 기반 분류 모델을 활용하여 국내 주식시장에서 급등주 포트폴리오의 수익률 특성을 실증적으로 분석하는 데 목적을 둔다. 급등주는 단기간에 비정상적인 수익을 기록하는 종목으로 정의되며, 본 연구에서는 하루 뒤 수익률이 20% 이상인 종목을 급등주로 규정하였다. 급등주는 단기적으로 높은 기대 수익을 제공하는 동시에 급격한 가격 변동과 낮은 예측 가능성으로 인해 투자자에게 높은 위험을 수반한다. 따라서 이러한 종목을 효과적으로 예측하고, 이를 활용한 포트폴리오 전략을 수립하는 것은 학문적·실무적으로 모두 의의가 크다.

효율적 시장 가설(Efficient Market Hypothesis, EMH)에 따르면, 모든 이용 가능한 정보는 즉각적으로 주가에 반영되므로 과거 데이터를 기반으로 초과수익을 예측하는 것은 불가능하다. 그러나 최근 인공지능과 빅데이터 분석 기법의 발전은 기존 가설의 한계에 도전할 가능성을 제시하고 있다. 본 연구는 2024 년 1 월부터 2025 년 4 월까지 KOSPI 와 KOSDAQ 종목을 대상으로 딥러닝 기반 급등주 예측 모델을 구축하고, 이를 활용하여 구성된 포트폴리오의 수익성을 검증함으로써 효율적 시장 가설에 대한 재검토의 근거를 제시한다.

특히 기존의 머신러닝, 딥러닝 기반 주가 예측 연구들은 대체로 수익률 예측에 초점을 두어, 포트폴리오 기반의 분석은 거의 이루어지지 않았다. 딥러닝 모델이 단순히 급등주를 예측하는 것을 넘어, 포트폴리오 관점에서 성능을 평가하는 것은 새로운 시사점을 제공할 여지가 있다.

II. 본론

2-1. 데이터 및 전처리

연구 데이터는 FnGuide 에서 제공하는 KOSPI 및 KOSDAQ 상장 종목의 일별 거래 정보를 활용하였다. 분석 대상 기간은 2024 년 1 월부터 2025 년 4 월까지이며, 이 중 2024 년 자료는 모델 훈련과 검증에 사용하고, 2025 년 1 월부터 4 월까지 자료는 모델 평가에 사용하였다. 금융업과 보험업 종목은 제외하였다. 2024 년 동안 총 239 거래일이 존재하며, 정상적으로 거래된 기업 수는 2,561 개였다. 하루 수익률이 20% 이상인 사례는 총 2,003 건으로 전체 관측치의 0.339%에 불과하여, 급등주 예측 문제는 심각한 불균형 분류 문제임을 확인할 수 있다.

이를 보완하기 위해 손실 함수에 pos_weight 옵션을 부여하여 급등주(양성 클래스)의 학습 비중을 강화하였다. 데이터는 훈련(train)과 검증(validation) 세트로 무작위 분할하였으며, 각각 80%와 20% 비율로 구성하였다. 입력 변수는 다양한 주식 시장 정보를 이용하여 15 개로 구성하여 모델에 입력하였다.

2-2. 모델 구조 및 학습 과정

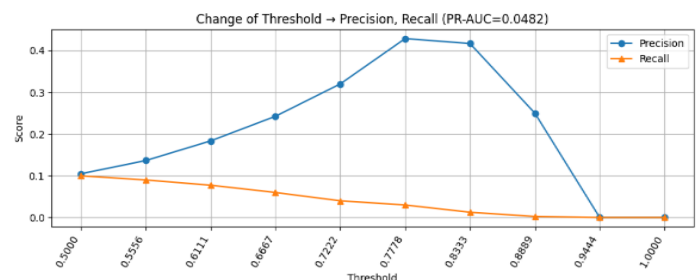
본 연구에서 활용한 딥러닝 모델은 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 구조로, 은닉층 차원은 [125, 25, 5]로 설정하였다. 각 은닉층은 Rectified Linear Unit(ReLU) 활성화 함수를 사용하였으며, 출력층은 이진 분류를 위해 시그모이드(sigmoid) 함수를 적용하였다. 최적화 알고리즘은 Adam 을 채택하였고, 학습률은 0.001 로 설정하였다. 배치 크기는 하나의 배치당 최소 하나의 급등주 데이터가 포함되는 규모로 훈련용 1,500 개, 검증용 400 개로 구성하였으며, 총 30 epoch 동안 학습을 진행하였다.

모델의 성능 평가 지표로는 Precision-Recall Area Under Curve(PR-AUC)를 채택하였다. 급등주와 같이 극히 드문 이벤트를 예측하는 문제에서는 정확도 지표가 무의미하고, 예측의 정확도를 나타내는 Precision 은 임계값과 양(+)의 상관관계를 지니기 때문에, 임계값 변화에 독립적인 PR-AUC 가 분류 성능을 비교하는 데 가장 적합하다. [표 1]은 다양한 하이퍼파라미터 조합에서 계산된 Precision 과 PR-AUC 값을 보여주며, 이 과정을 통해 본 연구의 모델을 최적의 조합으로 선정하였다.

[표 1] 파라미터 조합	Precision	PR-AUC
은닉층/은닉노드=[125, 25, 5] pos_weight_cap=22	0.089	0.0417
은닉층/은닉노드=[125, 25, 5] pos_weight_cap=21	0.0911	0.0398
은닉층/은닉노드=[125, 25, 5] pos_weight_cap=27	0.0945	0.0396

2-3. 임계값 조정 및 Trad-off 분석

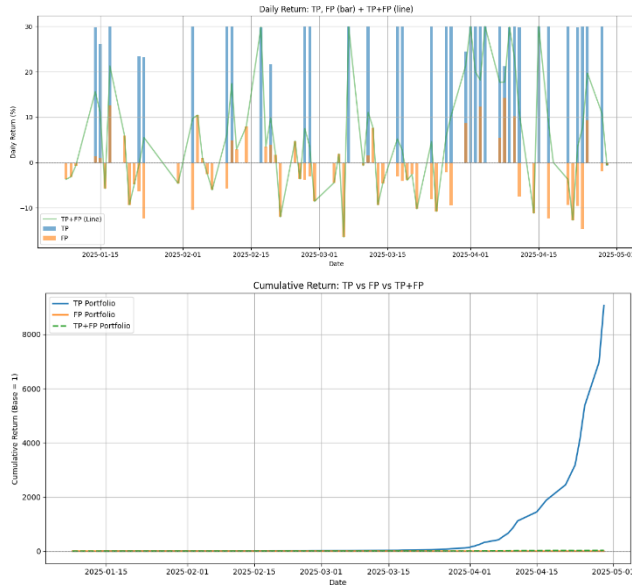
후보군으로 선정된 모델 조합에 대해, 예측 확률 임계값을 0~1 범위에서 100 단계로 조정하여 정밀도와 재현율 간 균형을 분석하였다. 연구 목적은 급등주 포트폴리오 수익 극대화에 있으므로, 재현율보다 정밀도를 우선하는 전략을 택하였다. 정밀도를 높이는 방향으로 모델을 개선해나가는 방법은 예측이 틀려 손실을 기록할 여지를 최대한 줄이기 위함이다. [그림 1]는 정밀도와 재현율의 변화 추이를 시각화한 결과로, 정밀도가 일정 수준 이상 확보되는 구간을 중심으로 최종 임계값을 결정하였다. 이 과정에서 모델은 FP(False Positive)를 줄이는 대신 TP(True Positive)를 집중적으로 탐지하는 방향으로 최적화되었다.



[그림 1] 임계값 변화에 따른 정밀도(파란색), 재현율(주황색)간 균형

2-4. 포트폴리오 성과 분석

최종 모델은 2025 년 1 월~4 월 데이터를 이용하여 평가되었으며, 포트폴리오는 TP, FP, TP+FP 세 가지 유형으로 구성되었다. [그림 3]과 [그림 4]는 각각 포트폴리오의 일별 수익률과 누적 수익률 추이를 보여준다.



TP 포트폴리오는 총 50 개 종목으로 평균 27.42%의 수익률을 기록하였다. 해당 종목들의 평균 주가는 약 8,052 원으로, 상대적으로 저가 종목에서 급등주가 빈번하게 발견됨을 의미한다. FP 포트폴리오는 145 개 종목으로 구성되었으며, 평균 주가 5,346 원으로 TP 와 유사하게 저가주 중심이었다. 이를 통해 모델이 주가 정보를 급등주 예측의 중요한 정보로 사용하고 있음이 드러난다. 일부 FP 종목은 양의 수익을 기록하기도 하나, 평가 기간 후반부에는 FP 포트폴리오의 누적수익률이 꾸준히 원금 이하를 기록하며 손실을 나타내었다. 따라서 FP 포트폴리오는 전체 포트폴리오 리스크 요인으로 작용한다고 평가하였다.

흥미로운 점은 전체 포트폴리오(TP+FP)가 지속적인 누적 수익을 기록했다는 사실이다. 이는 FP 의 손실에도 불구하고 TP 의 높은 수익이 전체 성과를 압도했기 때문이다. 다시 말해, 개별 예측 정확도의 한계에도 불구하고, 딥러닝 기반 전략은 포트폴리오 차원에서 실질적인 성과를 달성할 수 있음을 보여준다.

III. 결론

본 연구는 딥러닝 모델을 활용하여 급등주를 예측하고, 이를 바탕으로 포트폴리오 전략을 수립하여 성과를 검증하였다. 분석 결과, 모델의 예측력이 완벽하지는 않았으나, TP 포트폴리오의 높은 수익률이 FP 의 손실을 상쇄하여 전체 포트폴리오가 긍정적인 성과를 기록하였다. 이는 딥러닝 기법이 주식시장에서 특정 이벤트 기반 포트폴리오 전략을 설계하는 데 활용 가능성이 있음을 시사한다.

학술적으로 본 연구는 효율적 시장 가설에 대한 반론을 제기할 근거를 제공한다. 과거 주가와 거래 정보를 통해 미래 급등주를 일정 수준 예측할 수 있다는 점은, 정보가 즉각적이고 완전히 반영된다는 기존 가설을 재검토할 필요성을 보여준다. 실무적으로는 개인투자자 및 기관투자자 모두에게 고수익·고위험 전략으로서 급등주 예측 모형의 활용 가능성을 제시한다.

다만 본 연구는 데이터 기간이 제한적이고, 거래비용·세금·유동성 제약을 반영하지 못한 한계가 존재한다. 후속 연구에서는 장기간 데이터를 활용하여 시장 상황별 차이를 검토하고, 다양한 딥러닝 아키텍처(LSTM, Transformer 등)와의 비교 분석을 통해 모델 성능을 고도화할 필요가 있다. 또한 리스크 조정 수익률을 고려한 실질적 투자 성과 평가가 요구된다. 이러한 보완 작업을 통해 본 연구의 결과는 학술적·실무적으로 더욱 설득력을 가질 수 있을 것이다.

참 고 문 헌

- [1] Li, X., & Luo, C. (2020). An intelligent stock trading decision support system based on rough cognitive reasoning. *Expert Systems with Applications*, 160, 113763.
- [2] Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164, 163-173.
- [3] Kelly, B., Malamud, S., & Zhou, K. (2024). The virtue of complexity in return prediction. *The Journal of Finance*, 79(1), 459-503.
- [4] Gu, S., Kelly, B. T., & Xiu, D. (2019). Empirical asset pricing via machine learning. *Chicago Booth Research Paper*, (18-04), 2018-09.
- [5] Jin, S., Zhou, C., & Peng, H. (2023). Does price limit reduce stock price volatility on the limit up and down day?. *Finance Research Letters*, 58, 104302.
- [6] Li, H., Zheng, D., & Chen, J. (2014). Effectiveness, cause and impact of price limit—Evidence from China's cross-listed stocks. *Journal of International Financial Markets, Institutions and Money*, 29, 217-241.
- [7] Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455-1508.
- [8] Bauguess, S. W. (2019). Can Machine Learning in Finance Inform Clinical Decision Support?.