

포털 검색 강도와 딥러닝을 활용한 COVID-19 팬데믹 주가 예측

김 선 응

국민대학교 비즈니스IT전문대학원

Forecasting COVID-19 Pandemic Stock Prices using Portal Search Intensity and Deep Learning

Sun Woong Kim

Professor, Graduate School of Business IT, Kookmin University, Seoul 02707, Korea

[요 약]

네이버 트렌드에서 제공하는 키워드 검색지수는 그때그때의 사회적 이슈를 잘 반영하고 있다. 본 연구의 목적은 COVID-19 관련 네이버 검색어 지수와 장기 의존성 특성을 보이는 주가의 예측에서 우수한 성과를 보이는 딥러닝을 활용하여 주가를 예측하고자 한다. 실증 분석 결과는 다음과 같다. 첫째, 네이버의 코로나 검색 강도는 코스피 주가지수에 그랜저 인과성을 보여주었다. 둘째, 코스피 주가지수와 코스닥 주가지수 모두 LSTM 모형의 예측 성과가 단순 RNN 모형보다 우수한 예측 결과를 보여주었다. 셋째, LSTM 예측 주가를 이용한 알렉산더 필터 룰의 수익성이 Buy & Hold 전략의 수익성보다 높게 나타났다. 본 연구는 COVID-19 팬데믹 국면의 국내 주가 예측에서 네이버 검색지수의 유용성을 밝힌 첫 시도라는 점에서 학술적 의의가 크지만 다양한 딥러닝 모형으로 확장하여 분석하지 못한 점은 한계점이다.

[Abstract]

The keyword search index of Naver Trend well reflects current social issues. The purpose of this study is to predict stock prices using deep learning, which shows excellent performance in predicting stock prices showing long-term dependence with the Naver search index related to COVID-19. The results of the empirical analysis are as follows. First, Naver's Corona search intensity showed Granger causality in the KOSPI index. Second, LSTM model on both KOSPI and KOSDAQ indexes showed better predictive results than the simple RNN model. Third, the profitability of the Alexander Filter Rule using the LSTM predicted stock price was higher than that of the Buy & Hold strategy. This study is of great academic significance in that it is the first attempt to reveal the usefulness of the Naver search index in predicting domestic stock prices in the COVID-19 pandemic phase, but its failure to expand and analyze various deep learning models is a limitation.

색인어 : 코로나 팬데믹, 네이버 검색지수, 딥러닝 모형, LSTM, 알렉산더 필터 룰

Keyword : COVID-19 pandemic, Naver search index, Deep learning model, LSTM, Alexander filter rule

<http://dx.doi.org/10.9728/dcs.2022.23.2.343>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 30 December 2021; **Revised** 17 January 2022

Accepted 17 January 2022

***Corresponding Author; Sun Woong Kim**

Tel: +82-2-910-5471

E-mail: swkim@kookmin.ac.kr

I. 서론

Google이나 네이버와 같은 포털 검색 엔진 자료는 사회의 특정 이슈에 대한 대중들의 관심이나 걱정의 정도를 시의적절하게 잘 반영하고 있다. 전염병의 확산이나 주식시장의 이슈 발생 시 포털에서 특정 단어나 이슈의 검색량이 증가하기 때문에 이를 반영하면 전염병의 확산이나 주가의 방향성을 예측해 볼 수 있다. 특히 2020년 초 발생한 COVID-19의 주식시장 영향력은 어떤 전염병보다도 가장 강력하게 나타나고 있다[1], [2]. Lyocsa et al.(2020)은 Google Trends의 COVID-19 관련 검색량 정보를 이용하여 주요 10개국의 주가지수를 예측한 결과 우수한 예측 성과를 보여주었다[3]. Vasileiou(2021)는 COVID-19 확진자 증가 정보와 Google 검색 강도 정보를 이용하여 선진 10개국의 주가 영향력을 분석한 결과, COVID-19 확진자 증감 정보는 유의적인 영향력이 나타나지 않지만, Google 검색 강도는 주가와 유의적인 관련성이 있음을 밝혔다[4].

국내에서도 2020년 초 COVID-19 확진자가 발생함에 따라 투자자들 사이에 공포감이 확산하며 주가도 큰 폭의 하락을 경험하였고, 주가와 COVID-19 관련 연구도 활발히 진행되고 있다[5], [6]. Kim(2021)은 국내 COVID-19 일별 신규 확진자 발생 정보를 이용한 투자자 공포지수를 제안하고 자가와의 관련성을 분석한 결과 코스피 주가지수와 유의적인 그랜저 인과관계가 존재함을 밝혔다[7].

Tantaopas et al.(2016)은 ‘정보발견가설(information discovery hypothesis)’을 제안하고 투자자들이 새로운 뉴스에 대한 관심이 증가하는 경우, 해당 정보가 빠른 속도로 주가에 반영되면서 주가 변동성이 증가한다고 주장하였다[8]. Smales(2021)는 코로나 바이러스 관련 구글 검색량(Google search volume)을 투자자 관심도(investor attention)로 정의하고, 코로나 국면에서 주요 19개국 주가를 분석한 결과, 투자자 관심도가 주가 수익률에 (-)의 영향이 있음을 밝혀 정보발견가설을 지지하고 있다[9].

본 연구는 네이버 “검색어트렌드”에서 추출한 COVID-19 팬데믹 관련 검색어의 검색 강도(search intensity)를 활용하여 국내 주가지수를 예측하고자 한다. 네이버 검색 강도와 국내 주가지수 움직임 사이의 연관성을 분석하고, 네이버 검색 강도와 딥러닝 모형을 활용하여 미래 주가지수를 예측하고 실증 분석을 통해 예측 성과를 분석하고자 한다. 분석 자료는 COVID-19 관련 네이버 검색 강도, 코스피 주가지수, 코스닥 주가지수의 일별 자료이다. 본 연구는 COVID-19 팬데믹 국면에서 국내의 대표적인 포털 사이트인 네이버의 검색어 지수에 반영되는 투자자들의 불안 심리와 주가와 연관성을 분석한 첫 시도라는 점에서 학술적 의의가 있다.

II. 분석 모형

2-1 네이버 키워드검색 강도

Google이나 네이버 등의 포털 사이트는 검색 키워드에 대한 검색 빈도를 표준화한 검색량 지수를 제공하고 있다. 2020년 초 국내에서도 COVID-19 확진자가 발생함에 따라 사회적 불안감이 증가하기 시작하였고 이에 따라 COVID-19 관련 포털 사이트의 검색량도 증가하기 시작하였다. 2월 말 COVID-19 확진자가 급증하기 시작하면서 공포감은 극에 달하였다. 네이버 검색량 지수도 2020년 2월 말 급등하며 팬데믹의 불안 심리를 잘 반영하였다.

본 연구에서는 COVID-19 팬데믹에 대한 투자자들의 관심과 걱정의 정도를 측정하기 위하여 네이버 트렌드(Naver trend)를 이용하였다. 네이버 트렌드는 분석 기간의 검색 페이지에 대한 일별 검색지수에 대한 시계열 자료를 제공하고 있다. COVID-19 관련 네이버 검색지수(Naver search index)는 COVID-19 검색어로 검색되는 일별 검색 강도 지표로서, 분석 기간의 최고 검색량 값을 100으로 하는 상대적 표준화 비율이다. COVID-19 관련 검색어는 ‘코로나’, ‘코로나바이러스’, ‘COVID-19’, ‘COVID’, ‘팬데믹’을 포함하여 검색하였다.

COVID-19 확진자의 잠복기와 휴일에도 발표되는 네이버 검색지수를 평활화하기 위하여 네이버 검색지수를 바탕으로 식 (1)과 같은 코로나 검색 강도 CSI(COVID-19 search intensity) 지표를 제안하고 이를 이용하여 주가와 연관성을 분석한다.

$$CSI_t = \frac{\sum_{i=0}^2 Naver\ search\ index_{t-i}}{3} \div \frac{\sum_{i=0}^{13} Naver\ search\ index_{t-i}}{14}, \quad (1)$$

CSI_t is COVID-19 Search Intensity on day t.

식 (1)에서 분모는 코로나 잠복기가 최대 14일로 판단한 WHO의 견해에 따라 네이버 검색지수에 영향을 줄 수 있는 14일 동안의 네이버 검색지수 평균을 기준으로 하였으며, 분자는 주식시장이 폐장되는 주말에 발표되는 네이버 검색지수를 평활화하기 위하여 최근 3일의 네이버 검색지수 평균에 대한 상대적 크기로 CSI 지수를 산출하였다.

2-2 그랜저 인과관계를 이용한 네이버 검색지수의 주가 영향력 분석

COVID-19 팬데믹의 확산에 따라 COVID-19 관련 네이버 검색량도 심한 변동을 보여주었다. COVID-19 검색지수가 주식시장의 주가 변동과 관련성이 있는지를 그랜저 인과관계 모형(Granger causality model)을 이용하여 분석한다. 주가 수익률은 식 (2)와 같이 계산한다.

$$r_{i,t} = \frac{C_{i,t} - C_{i,t-1}}{C_{i,t-1}}, \quad (2)$$

$r_{i,t}$: return on stock index i on day t
 $C_{i,t}$: dose price of stock index i on day t

식 (2)에서 주가지수 i 의 일별 수익률(daily return)은 전 일 주가지수 대비 당일의 증가 변동량에 대한 상대적 비율로 측정한다.

그랜저 인과관계 모형은 식 (3)과 같다.

$$r_{i,t} = \omega_0 + \sum_{k=1}^p \alpha_{i,t} r_{i,t-k} + \sum_{k=1}^p \beta_{i,t} \Delta CSI_{t-k} + \epsilon_t, \quad (3)$$

$$\Delta CSI_{t-k} = CSI_t - CSI_{t-k}$$

$$H_0 : \beta_{i,t-1} = \beta_{i,t-2} = \dots = \beta_{i,t-p} = 0$$

식 (3)은 주가지수 i 의 수익률과 CSI의 변동량 사이의 p 시차 벡터 자기회귀모형이다. 그랜저 인과관계 모형의 귀무가설 H_0 은 CSI 변동량의 계수가 모두 0이며, 자료의 실증 분석 결과 귀무가설을 기각한다면 네이버 검색지수가 COVID-19 팬데믹 기간 주가와의 연관성이 있으며 이를 이용하면 예측의 정보효과가 존재함을 의미한다.

2-3 코로나 검색지수와 시계열 딥러닝 모형을 이용한 주가 예측

코로나 검색 강도를 이용하여 미래 주가를 예측하기 위하여 딥러닝 모형 중 시계열 자료의 예측에서 우수한 성과를 보여주는 신경망 모형의 하나인 RNN(recurrent neural network)을 활용하여 주가를 예측한다.

RNN은 데이터의 시간적 앞뒤 순서를 고려하지 않는 일반적인 신경망 모형과 달리 데이터의 시간 개념을 도입하여 시스템 내부적으로 입력변수를 기억함으로써 시계열 데이터의 시간적 순서를 고려할 수 있다. 기본 RNN 모형은 데이터의 시차가 길어질수록 과거의 데이터가 현재 시점에서 제대로 학습되지 못하는 기울기 소실 문제가 발생한다[10]. 주가와 같이 과거 데이터의 영향력이 장기 존재하는 경우는 주가의 예측에서 데이터의 장기 의존성(long-term dependence) 문제가 발생한다.

LSTM(long short-term memory) 모형은 이러한 장기 의존성 문제를 해결할 수 있도록 발전한 RNN 모형이다[11]. LSTM 모형은 기본 RNN 모형에 기존의 정보에 대하여 어느 정도로 다음 단계로 전달할 것인가를 결정하는 cell state를 추가하여 과거 정보도 현재 정보에 반영되도록 함으로써 장기 의존성 문제를 해결하고 있다.

초기 주가 연구에서는 인공신경망모형(artificial neural networks)을 이용한 주가 예측이 활발하였다[12]. Dingli and Fournier(2017)는 CNN(Convolutional Neural Networks) 모형을 이용하여 주가를 예측한 결과 주별 데이

터에서 60%의 예측 정확성을 보여주었지만 SVM 모형보다 예측 결과가 낮게 나타났다[13]. Moghar and Hamiche(2020)는 미국 주식 데이터를 이용한 LSTM 모형의 주가 예측 결과를 분석하고 epoch의 최적화의 필요성을 주장하였다[14]. 주가와 같은 시계열 데이터의 예측에서 LSTM 모형의 유용성은 여러 연구에서 밝혀지고 있다[15], [16]. Jing et al.(2021)은 주식 토론방 등에서 구한 온라인 정보로부터 CNN 모형을 통해 투자자 감성지수를 산출하고 LSTM 모형을 이용하여 중국 주가를 예측한 결과 예측력이 증가함을 보였다[17].

III. 데이터와 실험 설계

3-1 데이터 소개

네이버 트렌드 검색지수는 datalab.naver.com에서 구하였으며, COVID-19 키워드 검색지수를 바탕으로 코로나 검색지수인 CSI를 산출하였다. 분석 기간은 코로나 관련 네이버 검색어의 검색량이 본격적으로 나타나기 시작하는 2020년 1월 14일부터 2021년 11월 30일까지의 기간이다. 분석 기간에서의 주가 분석을 위해 한국거래소(Korea Exchange) 홈페이지에서 코스피 주가지수(KOSPI index), 코스닥 주가지수(KOSDAQ index)의 일별 자료를 구하였다.

코로나 관련 네이버 검색지수는 COVID-19 신규 확진자 증감의 영향을 받기 때문에 비교를 위해 COVID-19 신규 확진자 정보를 ourworldindata.org에서 구하였다. 그림 1은 분석 기간에서 COVID-19 신규 확진자와 코로나 관련 네이버 검색지수의 추이를 보여주고 있다. 한편, 그림 2는 자료의 분석 기간에서 코스피 주가지수와 코스닥 주가지수의 움직임을 보여주고 있다.

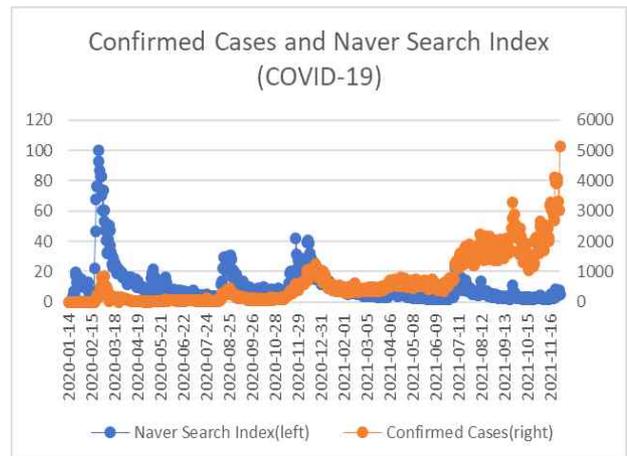


그림 1. 코로나 확진자와 네이버 검색지수 추이

Fig. 1. Trends of COVID-19 Confirmed Cases and Naver Search Index

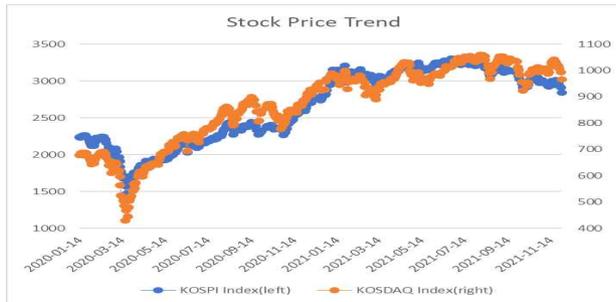


그림 2. 코스피 주가지수와 코스닥 주가지수 시계열 추이
Fig. 2. Time Series Data of KOSPI and KOSDAQ Index

그림 1에서 자료의 분석 기간 COVID-19 신규 확진자 수는 증감을 거듭하면서 2021년 하반기 이후에는 하루 확진자가 5,000명 이상 발생하면서 폭발적으로 증가하고 있다. 그러나 네이버 검색지수는 시간이 흐르면서 오히려 감소하고 있다. 네이버 검색지수는 2020년 초 COVID-19 신규 확진자가 발생하면서 코로나에 대한 불안 심리가 본격화되기 시작한 2020년 2월에 코로나 관련 검색량이 폭증하면서 검색지수가 2월 25일 정점을 기록하였다. 이후에는 확진자의 증감에 따라 소폭의 등락을 거듭하고 있다.

그림 2에서 자료의 분석 기간 주가지수는 코로나 확진자 발생에 따라 투자심리가 얼어붙으면서 하락을 거듭한 끝에 WHO의 코로나 팬데믹 선언을 전후하여 폭락세가 나타났고, 2020년 3월 19일 주식 투매로 주가지수가 최저점을 기록한 이후 지금까지 반등세가 이어지고 있다. 전체적으로 보면 코스피 주가지수와 코스닥 주가지수는 비슷한 패턴을 보여주고 있으며, COVID-19 확진자보다는 네이버 검색지수처럼 코로나 확산세 초반에 주가 하락국면이 나타났고 이어지는 구면에서는 주가는 상승국면이 이어지고 있다.

3-2 실험 설계

Smales(2021)는 COVID-19에 대한 글로벌 주식시장의 반응이 건강 위기(health crisis)에 대한 중요한 사례로 판단하고, 코로나 바이러스에 대한 투자자 관심도가 주가에 영향을 미치고 있음을 보여주었다[9]. 국내 주식시장에 대한 본 연구에서도 COVID-19 팬데믹 상황에서 네이버 검색지수가 국내 주가 수익률에 대한 그랜저 인과성이 존재하는지를 밝히고, 미래 주가의 예측 성과를 높이기 위하여 네이버 코로나 키워드 검색지수를 입력변수로 하는 주가 예측 모델을 제안한다. 본 연구의 실험 설계는 다음과 같이 2단계로 진행한다.

실험 1단계는 네이버 검색지수로부터 산출한 CSI와 주가 수익률 사이의 연관성을 분석한다. 분석 결과 네이버 검색지수에 기반한 CSI가 주가 수익률에 인과관계를 주고 있다면 2단계에서는 CSI와 과거 주가 자료를 입력변수로 하는 딥러닝 모형 LSTM을 통해 미래의 주가를 예측한다. 실험 2단계에서는 투자전략을 제안하고 예측된 주가에 기초하여 투자전략의 투자성적을 분석한다.

그림 3은 실험 2단계의 진행 과정을 보여주고 있다. 본 연구의 딥러닝 모형의 학습 데이터(training data)는 전체 기간의 70%에 해당하는 2020년 1월 14일부터 2021년 5월 10일까지, 검증용 데이터(test data)는 2021년 5월 11일부터 2021년 11월 30일까지의 기간이다. 본 연구는 Python 3.8.5 환경에서 Tensorflow 2.4.1과 Keras를 이용하여 실험을 진행하였다. 주가 예측을 위한 딥러닝 모형은 LSTM을 활용하며, 성과 비교를 위해 단순 RNN 모형도 적용한다. 딥러닝 모형의 최적화 알고리즘은 adam, 활성화 함수는 hyperbolic tangent, 지도학습은 최대 200회까지 실행하였다. 학습 데이터에 최적화된 LSTM 모형을 찾아 검증 데이터에서 예측 주가를 구하고 RMSE(root mean square error)를 이용하여 예측 성과를 비교하였다.

본 연구에서 활용할 투자전략은 주가의 상승과 하락에 따라 매수와 매도를 반복하는 대표적인 투자전략으로 알렉산더 필터 룰(AFR; Alexander filter rule)을 제안한다. AFR 투자전략은 필터의 크기에 따라 매수와 매도 시점이 결정된다. 구체적으로 x% AFR은 주가가 저점에서 x% 이상 상승하면 상승국면으로의 전환으로 인식하여 매수하며, 반대로 주가가 고점에서 x% 이상 하락하면 하락국면 전환으로 인식하여 매도하는 기계적인 거래전략(mechanical trading strategy)이다[18]. 식 (4)는 AFR 거래전략 식을 보여주고 있다.

Alexander Filter Rule (4)
 If close rises over x% from bottom then buy();
 If close falls down x% from top then sell();

k번째 거래에 따른 투자수익률(TR_k)은 식 (5)와 같다.

$$TR_k = \frac{C_{t2} - C_{t1}}{C_{t1}} \text{ for trade } k = 1, 2, \dots, n, \tag{5}$$

TR_k is k^{th} trading return,
 C_{t1} is the buy price, C_{t2} is the sell price.

분석 기간에서의 거래전략의 누적 수익률(CR; cumulative return)은 식 (6)과 같다.

$$CR = \prod_{k=1}^n (1 + TR_k) - 1 \tag{6}$$

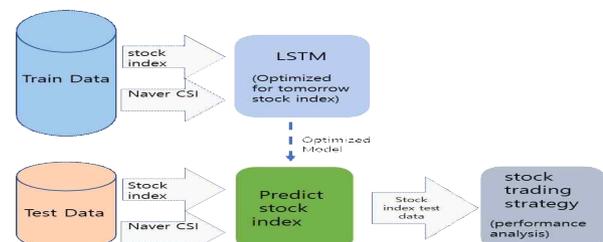


그림 3. 딥러닝 주가 예측 실험 설계
Fig. 3. Experimental Design on Deep Learning Stock Price Forecasting

IV. 실험 결과 분석

4-1 그랜저 인과관계 분석

COVID-19 팬데믹 구간에서의 COVID-19 관련 네이버 검색량 강도 지표인 CSI의 증가 수익률에 대한 인과관계를 그랜저 인과관계 식 (3)를 이용하여 실증 분석한다. 표 1은 증가 수익률의 정상성 검정(stationarity test)을 위한 ADF(augmented Dickey-Fuller test) 통계량과 그랜저 인과관계 F 통계량을 보여주고 있다.

표 1에서 증가 수익률 시계열이 안정적인지를 검증한 ADF 통계량은 코스피 증가지수의 경우 -13.11, 코스닥 증가지수의 경우 -10.27로 모두 1% 유의수준에서 ‘단위 근이 존재한다’는 귀무가설 H_0 을 기각하고 있다. 따라서 코스피 증가지수와 코스닥 증가지수 수익률은 모두 안정적인 시계열 데이터임을 보여주었다.

그랜저 인과관계 분석을 위한 식 (3)의 분석 결과 코스피 증가지수 수익률은 유의적인 결과를 보여주고 있어 Google Trends를 이용한 기존의 선진국 증가 연구 결과들과 일치하고 있다[19], [20]. 그러나 코스닥 증가지수 수익률은 귀무가설을 기각하지 못하고 있어 네이버 검색지수가 미래의 코스닥 증가지수 예측에서 유의미한 정보를 제공하지 못하고 있다.

4-2 딥러닝 예측 결과 분석

2020년 1월 14일부터 2021년 5월 10일까지의 학습 기간과 2021년 5월 11일부터 2021년 11월 30일까지의 검증 기간에서의 딥러닝 관련 제안모형인 RNN과 LSTM의 실험 결과는 표 2와 같다.

표 1. ADF 및 그랜저 인과관계 검정

Table 1. ADF and Granger Causality Tests

Stock Index	ADF	Granger Test F
KOSPI	-13.11 ***	6.69 ***
KOSDAQ	-10.27 ***	2.13

*** : significant at 1%

표 2. 딥러닝 예측모형의 RMSE

Table 2. RMSE of Deep Learning Forecasting Models

Period		RNN	LSTM
Training Period	KOSPI	0.046	0.018
	KOSDAQ	0.041	0.020
Test Period	KOSPI	0.059	0.020
	KOSDAQ	0.074	0.023

실험 2단계의 딥러닝 모형의 입력변수로는 코스피 증가지수와 코스닥 증가지수의 시가, 고가, 저가, 그리고 종가 수익률의 일별 자료이며, 이를 이용하여 다음 날의 증가를 예측한다. 표 2에서 코스피 증가지수는 학습 기간에서 RNN과 LSTM은 각각 0.046, 0.018, 검증 기간에서 RNN과 LSTM은 각각 0.059, 0.020의 RMSE 값을 보여주어 LSTM 모형의 예측 성과가 우수하게 나타났으며, 코스닥 증가지수 역시 학습 기간에서 RNN과 LSTM은 각각 0.041, 0.020, 검증 기간에서 RNN과 LSTM은 각각 0.074, 0.023의 RMSE 값을 보여주어 코스피 증가지수와 같은 결과가 나타났다. 전체적으로 LSTM의 예측 성능이 단순 RNN 모형보다 높은 결과를 보여주고 있어 증가의 장기 의존성이 영향을 미치고 있음을 알 수 있다. 코스닥 증가지수보다는 코스피 증가지수의 예측 성과가 더 높게 나타나고 있다.

그림 4와 그림 5는 검증 기간에서 LSTM 모형을 이용하여 코스피 증가지수와 코스닥 증가지수의 다음 날의 증가지수를 예측한 결과를 실제 증가와 비교하여 보여주고 있다.



그림 4. LSTM 모형의 예측 결과(코스피 증가지수)
Fig. 4. Forecasting Result on LSTM Model(KOSPI Index)

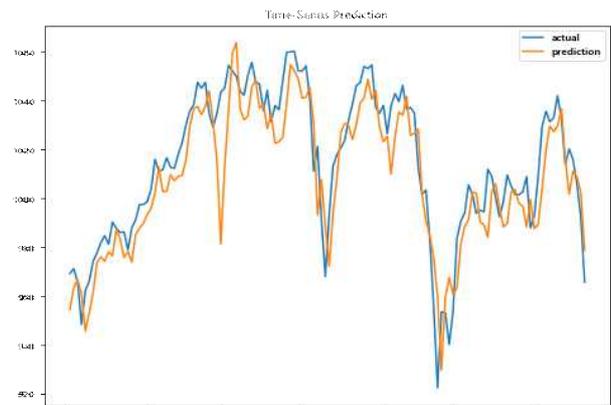


그림 5. LSTM 모형의 예측 결과(코스닥 증가지수)
Fig. 5. Forecasting Result on LSTM Model(KOSDAQ Index)

자료의 검증 기간에서 높은 예측 성과를 보여주고 있는 LSTM 모형의 예측 증가지수를 이용하여 제안된 알렉산더 필터 전략 AFR의 투자성과를 분석한다. 표 3과 표 4는 검증 기간에서 각각 코스피 시장과 코스닥 시장에 대한 AFR 거래 전략의 투자성과를 보여주고 있다. $Close^P$ 는 LSTM 모형에 의한 예측 증가를 이용한 투자성과, $Close^A$ 는 실제 증가지수를 이용한 투자성과를 비교하여 보여주고 있다. 검증 기간 코스피 주가지수는 -10.53% 하락하였고, 코스닥 주가지수는 -0.36% 하락하였다.

표 3. 알렉산더 필터 룰 누적 수익률(코스피 주가지수)

Table 3. Cumulative Returns on Alexander Filter Rule(KOSPI Index)

Alexander Filter	$Close^P$	$Close^A$
1%	-11.09%	-1.93%
2%	2.05%	-3.65%
3%	-0.46%	-8.38%
4%	-1.07%	-7.91%
5%	5.03%	1.62%

표 4. 알렉산더 필터 룰 누적 수익률(코스닥 주가지수)

Table 4. Cumulative Returns on Alexander Filter Rule(KOSDAQ Index)

Alexander Filter	$Close^P$	$Close^A$
1%	0.14%	12.42%
2%	20.24%	14.81%
3%	15.58%	19.80%
4%	11.61%	6.24%
5%	-14.95%	-4.89%

표 3의 코스피 주가지수에서 LSTM 예측 증가를 이용한 AFR 거래전략의 성과는 1% 필터를 제외한 필터에서 모두 실제 증가를 이용한 거래전략의 성과보다 개선되고 있다. 개선 폭은 필터 크기 2%에서 가장 높게 나타났다. 표 4의 코스닥 주가지수에서 LSTM의 예측 증가를 이용한 AFR 거래전략의 성과는 2% 필터와 4% 필터에서만 LSTM 예측 증가를 이용한 거래전략의 성과가 높게 나타났고, 1%, 3%, 5% 필터에서는 오히려 예측 증가의 투자성과가 더 낮게 나타났다. LSTM의 예측 성과가 더 낮게 나타났던 코스닥 주가지수의 경우 LSTM 예측 증가를 이용한 AFR 거래전략 역시 실제 증가를 이용한 거래전략의 성과보다 유리하지 않게 나오고 있다.

LSTM 예측 증가를 이용한 AFR 거래전략은 필터 크기가 2%일 때 코스피 주가지수와 코스닥 주가지수 모두 투자성과가 가장 큰 폭으로 개선되었다. 그림 6과 그림 7은 코스피 주가지수와 코스닥 주가지수에 대하여 LSTM 예측 증가를 이용한 2% 필터 AFR 거래전략의 수익 곡선(equity curve)을 매수 후 단순 보유전략인 Buy & Hold 전략의 수익 곡선과 비교하여 보여주고 있다.

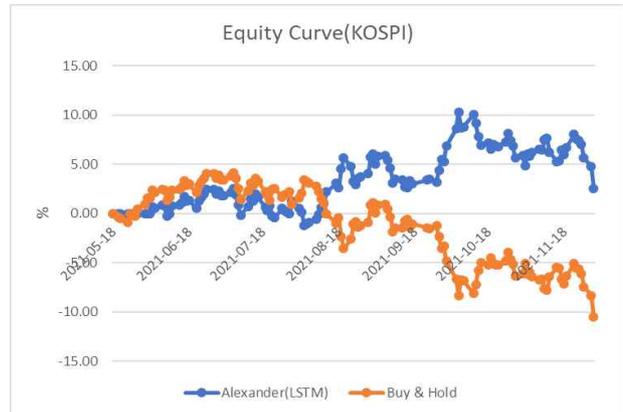


그림 6. LSTM AFR 거래전략의 수익 곡선(코스피 주가지수)
Fig. 6. Equity Curve on LSTM AFR(KOSPI Index)

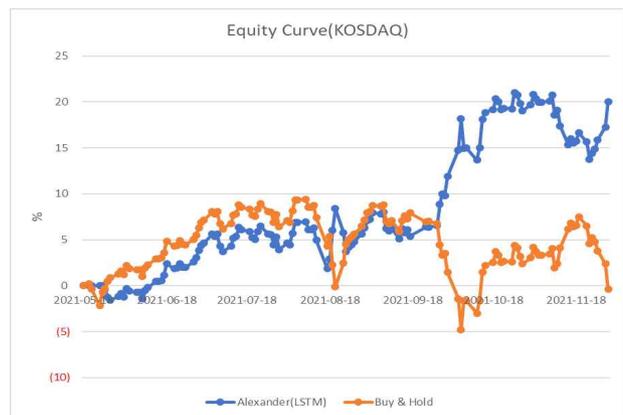


그림 7. LSTM AFR 거래전략의 수익 곡선(코스닥 주가지수)
Fig. 7. Equity Curve on LSTM AFR(KOSDAQ Index)

그림 6과 그림 7에서 모두 LSTM 예측 증가를 이용하는 알렉산더 필터 전략의 투자성과는 아무런 예측이나 고안된 거래전략 없는 Buy & Hold 전략보다 상당히 개선되는 수익 곡선을 보여주고 있다. 본 연구의 결과는 투자 실무적 측면에서 투자자들에게 새로운 투자 접근법을 제시하고 있다.

V. 결론

Google이나 네이버와 같은 포털에서 제공하기 시작한 특정 이슈에 대한 검색지수의 유용성은 여러 분야에서 나타나고 있다. 본 연구의 목적은 네이버 트렌드에서 제공하고 있는 COVID-19 관련 검색 키워드의 검색 강도를 이용하여 팬데믹 국면에서 증가를 예측하고 거래전략을 통해 투자전략의 투자 성과를 분석하였다. 주가 예측의 개선을 위해 딥러닝 모형 중 주가와 같은 장기 의존성 특성을 보이는 시계열 자료의 예측에서 우수한 예측 성과를 보이는 LSTM 모형을 활용하였다.

실증 분석 결과는 다음과 같다. 첫째, 네이버 검색지수에 기반하여 산출된 코로나 검색 강도는 코스피 주가지수에 유의적인 그랜저 인과관계를 보여주고 있지만, 코스닥 주가지수

에서는 유의적인 인과성이 나타나지 않았다. 둘째, 주가 시계열의 장기 의존성 특성을 반영하여 단순 RNN 모형보다 LSTM 모형의 예측 성과가 더 높게 나타났다. 셋째, 코스피 주가지수가 코스닥 주가지수보다 LSTM의 예측 성과가 높게 나타났다. 마지막으로, 제안된 알렉산더 필터 룰의 거래 성과는 코스피 주가지수와 코스닥 주가지수 모두 Buy & Hold 전략보다 우수한 수익 곡선을 보여주었다.

일반적으로 주가 대폭락과 같은 금융위기는 신용 위기나 경제의 본질가치 하락에 따른 위기 국면에서 발생한다. COVID-19 팬데믹 주가 대폭락은 경제 본질가치보다는 보건 위기에 의해서 촉발된 사례라는 점에서 과연 코로나 팬데믹과 같은 비경제적 요인이 주가에 영향을 미치는지를 분석한 본 연구는 비경제적 요인의 주가 연구에 대한 결과물을 하나 추가하고 있다. 특히, COVID-19 팬데믹 국면에서 네이버와 같은 포털 사이트의 COVID-19 관련 검색 강도를 이용하여 국내 주식시장에서 팬데믹 주가를 예측한 첫 시도라는 점에서 학술적 의의를 찾을 수 있다. COVID-19 유행병과 같은 비경제적 변수도 주가 예측에서 유용성이 있는 것으로 밝혀짐에 따라 실무적 관점에서는 투자자들에게 확장된 투자전략을 제시하였다.

본 연구는 다음과 같은 한계점을 가지고 있다. 첫째, RNN이나 LSTM 모형 이외의 다양한 딥러닝 모형과의 비교분석이 부족하다는 점이다. 둘째, 왜 코스닥 주가지수에서는 네이버 검색 강도가 영향력이 나타나지 않는지에 대한 분석이 부족하다는 점을 들 수 있다. 셋째, 코로나 관련 검색지수를 추출하는 과정에서 키워드의 중요성에 비추어보면 코로나 관련 새로운 키워드를 추가하여 산출된 검색지수에서의 강건성 테스트도 필요할 것으로 판단되며, 향후 연구에서 이러한 한계점들이 밝혀지기를 기대한다.

참고문헌

- [1] S. Baker, N. Bloom, S. Davis, K. Kost, M. Sammon, and T. Viratyosin, "The unexpected stock market impact of COVID-19," *NBER Working Paper*, No. 26945, pp. 1-24, April 2020. <https://doi.org/10.3386/w26945>
- [2] C. D. Garcia-Gomez, E. Demir, J. M. Diez-Esteban, and Y. Bilan, "The impact of COVID-19 outbreak on hotels' value compared to previous diseases: the role of ALFO strategy," *Heliyon*, Vol. 7, e07836, pp. 1-8, August 2021. <https://doi.org/10.1016/j.heliyon.2021.e07836>
- [3] S. Lyocsa, E. Baumohl, T. Vyrost, and P. Molnar, "Fear of the coronavirus and the stock markets," *Finance Research Letters*, Vol. 36, 101735, pp. 1-7, August 2020. <https://doi.org/10.1016/j.frl.2020.101735>
- [4] E. Vasileiou, "Explaining stock markets' performance during the COVID-19 crisis: Could Google searches be a significant behavioral indicators?" *Intelligent Systems in Accounting, Finance and Management*, Vol. 28, No. 3, pp. 173-181, July 2021. <https://doi.org/10.1002/isaf.1499>
- [5] Y. Cho and D. Kim, "The effect of COVID-19 for FSC, LCC, shipping and logistics company stock volatilities," *Journal of the Aviation Management Society of Korea*, Vol. 19, No. 5, pp. 37-54, October 2021. <https://doi.org/10.30529/amsok.2021.19.5.003>
- [6] S. W. Kim, "COVID-19 pandemic and investor herding behavior," *Journal of Digital Contents Society*, Vol. 22, No. 7, pp. 1083-1090, July 2021. <https://doi.org/10.9728/dcs.2021.22.7.1083>
- [7] S. W. Kim, "COVID-19 fear index and stock market," *Journal of Convergence for Information Technology*, Vol. 11, No. 9, pp. 84-93, September 2021. <https://doi.org/10.22156/CS4SMB.2021.11.09.084>
- [8] P. Tantaopas, C. Padungsaksawasdi, S. Treepongkaruna, "Attention effect via internet search intensity in Asia-Pacific stock markets," *Pacific-Basin Finance Journal*, Vol. 38, pp. 107-124, June 2016. <https://doi.org/10.1016/j.pacfin.2016.03.008>
- [9] L. A. Smales, "Investor attention and global market returns during the COVID-19 crisis," *International Review of Financial Analysis*, Vol. 73, pp. 1-14, January 2021. <https://doi.org/10.1016/j.irfa.2020.101616>
- [10] A. Yadav, C. K. Jha, and A. Sharan, "Optimizing LSTM for time series prediction in Indian stock market," *Procedia Computer Science*, Vol. 167, pp. 2091-2100, April 2020. <https://doi.org/10.1016/j.procs.2020.03.257>
- [11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, Vol. 9, No. 8, pp. 1735-1780, November 1997. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [12] F. A. de Oliveira, C. N. Nobre, and L. E. Zarate, "Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index - Case study of PETR4, Petrobras, Brazil," *Expert Systems with Applications*, Vol. 40, No. 18, pp. 7596-7606, December 2013. <https://doi.org/10.1016/j.eswa.2013.06.071>
- [13] A. Dingli and K. S. Fournier, "Financial time series forecasting - A deep learning approach," *International Journal of Machine Learning and Computing*, Vol. 7, No. 5, pp. 118-122, October 2017. <https://doi.org/10.18178/ijmlc.2017.7.5.632>
- [14] A. Moghar and M. Hamiche, "Stock market prediction using LSTM recurrent neural network," *Procedia Computer Science*, Vol. 170, pp. 1168-1173, April 2020. <https://doi.org/10.1016/j.procs.2020.03.049>

- [15] S. W. Kim, "Performance analysis of Bitcoin investment strategy using deep learning," *Journal of the Korea Convergence Society*, Vol. 12, No. 4, pp. 249-258, April 2021. <https://doi.org/10.15207/KCS.2021.12.4.249>
- [16] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, and A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis," *IEEE Access*, Vol. 8, pp. 226-241, August 2020. <https://doi.org/10.1109/ACCESS.2020.3015966>
- [17] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, Vol. 178, pp. 1-12, April 2021. <https://doi.org/10.1016/j.eswa.2021.115019>
- [18] E. F. Fama and M. E. Blume, "Filter rules and stock market trading," *The Journal of Business*, Vol. 39, No. 1, pp. 226-241, January 1966.
- [19] M. Costola, M. Iacopini, C. Santagiustina, "Google search volumes and the financial markets during the COVID-19 outbreak," *Finance Research Letters*, Vol. 42, 101884, pp. 1-7, December 2020. <https://doi.org/10.1016/j.frl.2020.101884>
- [20] A. K. Dey, G. M. Toufiqul Hoque, K. P. Das, and I. Panovska, "Impacts of COVID-19 local spread and Google search trend on the US stock market," *Physica A: Statistical Mechanics and Its Applications*, In press, September 2021. <https://doi.org/10.1016/j.physa.2021.126423>



김선웅(Sun Woong Kim)

1983년 : 한국과학기술원 경영과학과 (공학석사)

1988년 : 한국과학기술원 경영과학과 (공학박사-금융공학)

1988년~1998년: 보람투자신탁운용주식회사

1999년~2003년: 원업투자자문주식회사

2004년~2008년: 에프이테크주식회사

2009년~현 재: 국민대학교 비즈니스IT전문대학원 트레이딩시스템전공 주임교수

※관심분야 : 트레이딩시스템, 자산운용, 투자위험관리