

# 코스피 방향 예측을 위한 하이브리드 머신러닝 모델

황희수  
한라대학교 ICT공학부 교수

## Hybrid Machine Learning Model for Predicting the Direction of KOSPI Securities

Heesoo Hwang  
Professor, Division of Information Communication Technology, Halla University

**요약** 과거 주가 데이터와 금융 관련 빅 데이터를 사용해 머신러닝 기법으로 주식시장을 예측하는 연구는 다양하게 있어 왔지만, HTS와 MTS를 통해 거래가 가능한 주가지수 연동 ETF가 생기면서 주가지수를 예측하는 연구가 최근 주목받고 있다. 본 논문에서는 KOSPI 연동 ETF를 거래할 목적으로 KOSPI의 상승 예측을 위한 머신러닝 모델과 하락 예측을 위한 모델을 각각 구현한다. 이들 모델은 매개변수의 그리드 탐색을 통해 최적화 된다. 또한 정밀도를 개선해 ETF 거래 수익률을 높일 수 있도록 개별 모델들을 조합한 하이브리드 머신러닝 모델을 제안한다. 예측 모델의 성능은 정확도와 ETF 거래 수익률에 큰 영향을 미치는 정밀도로 평가된다. 하이브리드 상승 예측 모델의 정확도와 정밀도는 72.1 %와 63.8 %이고 하락 예측 모델은 79.8 %와 64.3 %이다. 하이브리드 하락 예측 모델에서 정밀도는 개별 모델보다 최소 14.3 %, 최대 20.5 % 개선되었다. 테스트 기간에 하이브리드 모델은 하락에서 10.49 %, 상승에서 25.91 %의 ETF 거래 수익률을 보였다. 인버스×2와 레버리지 ETF로 거래하면 수익률을 1.5 ~ 2배로 높일 수 있다. 하락 예측 머신러닝 모델에 대한 추가 연구로 수익률을 더 높일 수 있을 것으로 기대한다.

**주제어** : 융합, 주가 방향, 예측, ETF, 수익률, 기계학습, 하이브리드 모델

**Abstract** In the past, there have been various studies on predicting the stock market by machine learning techniques using stock price data and financial big data. As stock index ETFs that can be traded through HTS and MTS are created, research on predicting stock indices has recently attracted attention. In this paper, machine learning models for KOSPI's up and down predictions are implemented separately. These models are optimized through a grid search of their control parameters. In addition, a hybrid machine learning model that combines individual models is proposed to improve the precision and increase the ETF trading return. The performance of the prediction models is evaluated by the accuracy and the precision that determines the ETF trading return. The accuracy and precision of the hybrid up prediction model are 72.1 % and 63.8 %, and those of the down prediction model are 79.8% and 64.3%. The precision of the hybrid down prediction model is improved by at least 14.3 % and at most 20.5 %. The hybrid up and down prediction models show an ETF trading return of 10.49%, and 25.91%, respectively. Trading inverse×2 and leverage ETF can increase the return by 1.5 to 2 times. Further research on a down prediction machine learning model is expected to increase the rate of return.

**Key Words** : Convergence, Stock Movement, Prediction, ETF, Return, Machine Learning, Hybrid Model

\*Corresponding Author : Heesoo Hwang(hshwang@halla.ac.kr)

Received March 10, 2021

Accepted June 20, 2021

Revised April 5, 2021

Published June 28, 2021

## 1. 서론

정치, 사회, 경제 등의 다양한 변수에 영향을 받는 주식 시장에서 주가 예측은 쉽지 않은 것으로 여겨지고 있지만 그럼에도 주가 예측에 다양한 머신러닝 방법들이 적용되었고[1-5], 주가 대신 주가의 방향을 예측해 거래에 활용하려는 연구도 활발하다. 최적화된 신경회로망을 사용한 주가 예측[6], S&P 500 주식의 주가 방향 예측[7]과 주가의 예측 정확도를 개선하기 위한 머신닝 앙상블 기법[10-13] 및 주가와 뉴스 등의 온라인 데이터 소스를 학습해 주가를 예측하는 연구가 있다[14, 15]. 기존 연구가 주로 종목별 주가 예측을 목표로 하는 반면에, 최근에는 HTS, MTS 등을 통해 거래가 가능한 주가지수 연동 ETF가 생기면서 주가지수 예측을 위한 연구가 최근 주목받고 있다[1, 2, 8-9].

본 논문은 Korea composite Stock Price Index (KOSPI)에 연동한 ETF의 거래를 목적으로 KOSPI의 방향을 예측한다. 주가에 영향을 미치는 다양한 빅 데이터 대신에 오직 KOSPI의 과거 증가 데이터만을 사용하여 KOSPI의 일일 방향을 예측하는 머신러닝 모델을 제안한다. KOSPI의 상승 또는 하락 방향을 예측할 수 있도록 머신러닝 기법인 Multi-Layer Perceptrons Classifier (MLPC), Support Vector Classifier (SVC), Random Forest Classifier (RFC), Light Gradient Boosting Machine Classifier (LGBMC) 및 Extreme Gradient Boost Classifier (XGBC)를 사용해 상승 및 하락 예측 모델을 각각 구현한다. 이들 모델은 모델별 매개변수의 그리드 탐색을 통해 최적화된다. 구현된 모델의 성능은 예측 정확도와 ETF 거래 수익률에 큰 영향을 미치는 정밀도로 평가된다. 개별 모델을 조합해 예측 정밀도를 개선한 하이브리드 머신러닝 예측 모델을 제안하고 이를 토대로 ETF 거래를 시뮬레이션 한다. 예측 모델의 검증(verification)과 테스트를 위해 2018년과 2019년 ~ 2020년의 KOSPI 데이터를 사용하여 정확도, 정밀도 및 ETF 거래 수익률을 제시한다. 2장에서는 주가 방향 예측을 위한 입력력 변수 설정과 머신러닝 모델이, 3장에서는 머신러닝 모델을 조합한 하이브리드 모델이, 4장에서는 시뮬레이션 및 성능 평가가 다루어지고 5장에 결론이 기술된다.

## 2. 주가 방향 예측

### 2.1 입력과 출력

주가에 영향을 미치는 요인은 매우 다양하며 미래의 주가를 예측하는 방법도 기업의 본질적 가치를 따지는 기본 분석부터, 기술 분석, 효율적 시장 이론 등 다양하다[5]. 이중 기술 분석은 주가, 거래량 등 시장 내부 데이터만을 사용해 주가를 예측하는 방법이다. 이 분석에서는 주가지수, 거래량 등의 데이터 변화를 중시하고 주가의 특이한 변화 패턴을 알아내기 위해 차트 분석을 사용한다[17]. 기본 분석이 투자종목을 선택하는 데 유력하다면 기술 분석은 주식의 거래 시점을 선택하는 데 사용된다. 기술 분석에서 어떤 변수를 사용하는 것이 예측에 유리한지에 대한 객관적인 연구는 없다. 본 논문에서는 주식 시장 외부 요인은 배제하고 시장 내부 데이터 중 오직 주가의 변화만을 고려한다. 즉, 주가의 움직임이 과거와 유사한 패턴이 되면 주가는 과거와 유사하게 움직인다고 가정한다. 캔들스틱 차트의 패턴 연구에 의하면 주가 변화를 예측하는데 영향력이 있는 캔들스틱 패턴은 대부분 3일 이내의 것이다[17].

본 논문에서는 KOSPI의 일일 방향을 예측하기 위해 오직 과거 KOSPI 증가 데이터만을 사용한다. 예측에 사용할 과거 데이터를 결정하기 위해 자기 상관 분석을 수행한다. Fig. 1은 95% 신뢰도 구간에서 KOSPI 증가에 대한 과거 20일의 자기상관 분석이고 Fig. 2는 동일 조건에서 식(1)과 같이 KOSPI의 전일 대비 변동률에 로그를 취해 [-1,1]의 범위로 변환한 데이터의 자기 상관 분석이다.

$$inp_t = \log \frac{KOSPI_t}{KOSPI_{t-1}} \quad (1)$$

여기서  $KOSPI_t$ 는 KOSPI의 증가이고  $KOSPI_{t-1}$ 는  $KOSPI_t$ 의 1일 전 데이터이다.

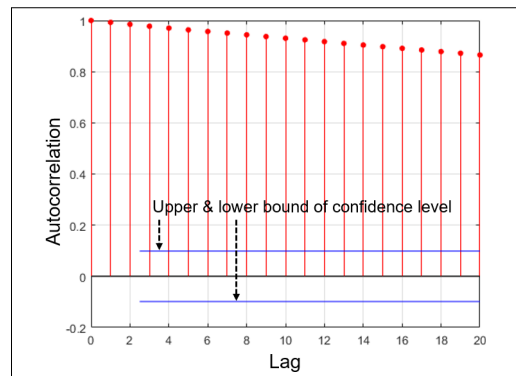


Fig. 1. Autocorrelation of  $KOSPI_t$

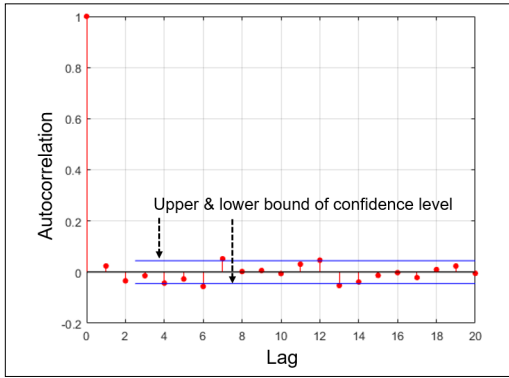


Fig. 2. Autocorrelation of  $\log \frac{KOSPI_t}{KOSPI_{t-1}}$

Fig. 1에서 과거 20일 데이터 모두가 자기상관 분석의 신뢰도 상한 위에 있지만, Fig. 2에서는 신뢰도 상한 및 하한을 넘는 데이터는  $inp_{t-3}$ ,  $inp_{t-5}$ ,  $inp_{t-6}$ ,  $inp_{t-11}$ 와  $inp_{t-12}$ 이다.  $KOSPI_{t+1}$ 의 방향 예측에 과거 8일 데이터로 만든  $inp_t$ ,  $inp_{t-1}$ ,  $inp_{t-2}$ ,  $inp_{t-3}$ ,  $inp_{t-4}$ ,  $inp_{t-5}$ 와  $inp_{t-6}$ 를 입력으로 사용한다.

예측하고자 하는 KOSPI의 방향은 상승 (Up) 또는 하락 (Down)으로, 식(2)의 변동률  $p_{t+1}$ 가  $\alpha$  (%) 이상이면 상승,  $-\alpha$  (%) 이하이면 하락으로 정의한다.

$$p_{t+1} = \frac{KOSPI_{(t+1)} - KOSPI_{(t)}}{KOSPI_{(t)}} \times 100 \quad (2)$$

### 2.2 예측 모델

주식 시장 예측에 머신러닝 기술을 적용하는 것은 예측 정확도가 개선된 효과적이고 효율적인 모델을 개발하기 위함이다. 이는 유익한 예측으로 높은 투자 수익을 기대할 수 있기 때문이다. 과거에 머신러닝 기술이 개별적으로 적용되었다면 최근에는 머신러닝 알고리즘을 다양하게 조합해서 예측 성능을 높이려는 시도가 활발하다 [10-11, 14-16]. 머신러닝에서 다수의 알고리즘을 조합해 사용하는 방식을 앙상블이라 한다. 앙상블 방식은 분산을 줄여서 예측 정확도를 개선할 수 있다.

본 논문에서는 KOSPI의 일일 방향을 예측하기 위해 단일 머신러닝 알고리즘을 적용하는 모델과 다수의 알고리즘을 적용하는 앙상블 방식의 모델을 구현한다. 예측 모델로 사용될 MLPC와 SVC는 단일 알고리즘 적용 방식이고 RFC, LGBMC 및 XGBC는 앙상블 방식이다. MLPC는 신경회로망의 우수한 패턴 분류 기능으로 주가 방향 예측에 널리 사용되어온 모델이다. SVC는 전통적

으로 고차원 데이터 공간에서 데이터의 비선형 분리에 효과적인 것으로 알려진 모델이다. RFC는 의사결정 트리 기반의 앙상블 알고리즘으로 대용량 데이터 처리에 효과적이며 과적합 문제를 회피하여 모델의 정확도를 향상시킨다. LGBMC는 트린 부분에 가중치를 더하면서 진행하는 앙상블 알고리즘으로, 기존의 트리 기반 알고리즘과 달리 수직적으로 확장하는 트리 구조를 가지며 메모리를 적게 차지하고 속도가 빠른 장점이 있다. XGBC는 다수의 의사결정 트리를 조합해서 사용하는 앙상블 알고리즘으로, 구조화된 데이터를 갖고 예측할 때 매우 우수한 성능을 보인다. 이와 같이 서로 다른 특징을 갖는 머신러닝 모델을 사용해서 Fig. 3에 보인 예측 모델을 구현한다. 다중입력-단일출력의 이 예측 모델은 KOSPI 상승 예측과 하락 예측에 대해 각각 구현되며, 모델의 출력은 식(3)과 식(4)로 기술된다.

$$out_{t+1}^{Up} = \begin{cases} 1 & \text{if } p_{t+1} \geq \alpha \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$out_{t+1}^{Down} = \begin{cases} 1 & \text{if } p_{t+1} \leq -\alpha \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

상승 및 하락 예측 모델의 출력이 각각 1과 0이면 상승을, 0과 1이면 하락을 나타내고 0과 0 또는 1과 1이면 예측은 무시된다.

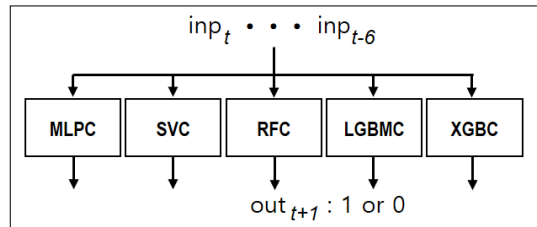


Fig. 3. Machine learning models for predicting the daily direction of KOSPI

### 3. 하이브리드 머신러닝 모델

KOSPI 방향 예측을 위해 구현된 개별 머신러닝 모델의 예측 결과를 조합해 사용하는 Fig. 4의 하이브리드 머신러닝 예측 모델을 제안한다. 개별 머신러닝 모델은 그 알고리즘이 서로 상이하고 장단점이 다르기 때문에 예측 결과(상승 또는 하락으로 예측된 날)도 다를 수 있다. 하이브리드 모델은 서로 다른 머신러닝 알고리즘 모델의 상승 또는 하락 예측 결과를 합해 사용함으로써 예측 정밀도를 개선해 결과적으로 KOSPI ETF 및 인버스 ETF의 거래 수

익률을 높일 수 있다. Fig. 4에 보인 것처럼 하이브리드 상승 예측 모델은 상승 예측 MLPC, SVC, RFC, LGBMC 및 XGBC의 출력이 합해지고 그 합이  $\beta$  이상이면 1로, 그렇지 않으면 0으로 출력한다. 마찬가지로, 하이브리드 하락 예측 모델도 하락 예측 MLPC, SVC, RFC, LGBMC 및 XGBC의 출력이 합해지고 그 합이  $\beta$  이상이면 1로, 그렇지 않으면 0으로 출력한다.

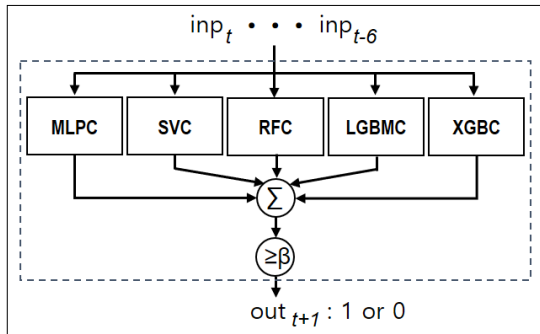


Fig. 4. Hybrid machine learning models for predicting the direction of KOSPI

## 4. 시뮬레이션

### 4.1 머신러닝 모델

KOSPI의 상승 및 하락 방향을 예측하기 위해 Fig. 3에 보인 5가지 모델을 각각 구현한다. 모델 학습 및 모델 검증에 각각 2012년 ~ 2017년과 2018년의 KOSPI 증가 데이터를 사용한다. 모델의 최종 테스트는 2019년 ~ 2020년의 KOSPI 데이터로 수행된다. Fig. 5와 6은 검증 데이터에서 KOSPI가 각각  $\alpha$  이상 오르거나  $-\alpha$  이하로 하락한 날을 보인다.  $\alpha$  값이 작으면 잡음 데이터에 대해서도 상승과 하락을 예측해야 되고 이는 예측의 정확도와 정밀도를 낮춘다[9]. 본 논문에서는  $\alpha = 0.7$ 을 사용하며, 이는 잡음성인 날을 배제시켜 예측 대상인 날을 전체의 20% 내외로 줄여준다.

머신러닝 예측 모델을 구현하는데 사용할 최적의 매개변수는 그리드 탐색을 통해 찾아진다. 모델별 매개변수의 그리드 탐색 범위는 Table 1과 같고, 탐색 결과 상승 예측 모델의 최적 매개변수는 다음과 같다. MLPC에서 iteration 430, learning\_rate 0.01, 은닉 층 2, 노드 수 11과 12이다. SVC에서 C 9.8, gamma 79이다. RFC에서 앙상블 알고리즘의 수 5, max\_depth 10,

min\_sample\_split 19, min\_sample\_leaf 5이다. LGBMC의 learning\_rate 0.25, max\_depth 2, leaf\_max\_data 191, feature\_fraction 0.5, leafs 3, bagging\_fraction 0.8이다. XGBC의 알고리즘 수 7, max\_depth 12, min\_child\_weight 2, gamma 0.9이다. 하락 예측 모델의 최적 매개변수는 다음과 같다. MLPC에서 iteration 695, learning\_rate 0.01, 은닉 층 2, 노드 수 15와 17이다. SVC에서 C 1.6, gamma 242이다. RFC에서 앙상블 알고리즘의 수 2, max\_depth 13, min\_sample\_split 23, min\_sample\_leaf 6이다. LGBMC의 learning\_rate 0.25, max\_depth 3, leaf\_max\_data 53, feature\_fraction 0.5, leafs 7, bagging\_fraction 0.6이다. XGBC의 앙상블 알고리즘 수 7, max\_depth 13, min\_child\_weight 1, gamma 0.1이다.

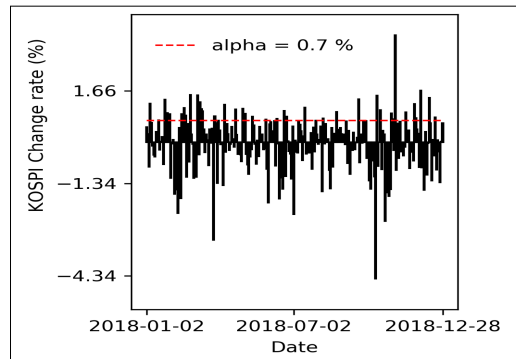


Fig. 5. Days when the KOSPI rose above 0.7% in 2018

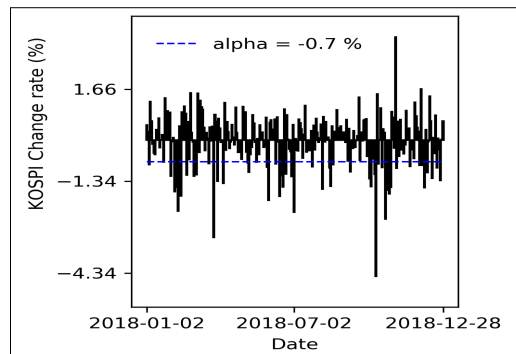


Fig. 6. Days when the KOSPI fell below -0.7% in 2018

**Table 1. Grid search ranges for optimal parameters of machine learning models**

Models	Parameters	Grid search ranges		
		Start	Stop	Increase
MLPC	iteration	300	900	5
	node	5	20	1
SVC	C	1	10	0.1
	gamma	50	300	1
RFC	algorithms	2	10	1
	max_depth	2	15	1
	min_sample_split	5	25	1
	min_sample_leaf	2	10	1
LGBMC	max_depth	2	15	1
	leaf_max_data	50	200	1
	leafs	2	10	1
	feature_fraction	0.4	0.6	0.1
	bagging_fraction	0.5	0.9	0.1
XGBC	algorithms	2	10	1
	max_depth	2	15	13
	min_child_weight	1	5	1
	gamma	0.1	0.9	0.1

Table 2과 3은 테스트 기간 동안 상승 및 하락 예측 모델의 정확도와 정밀도를 보인다. 상승 예측을 예로 들면, 식(5)의 정확도는 모든 상승 예측에 대한 옳은 상승 예측과 옳은 상승 아님 예측의 비율을 표시하며, 식(6)은 상승 예측 중에 옳은 예측의 비율을 표시한다. KOSPI ETF 거래에서 손실을 줄이려면 예측 정밀도가 높아야 한다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \times 100 \quad (6)$$

Table 2에서 상승 예측 모델 LGBMC의 정밀도가 낮았고 Table 3에서 하락 예측 모델의 정확도는 우수하지만 정밀도는 50 % 이하이다. 이는 하락 예측에 의한 인버스 ETF 거래에서 손실 발생 가능성을 높인다.

주식시장을 예측하는 연구는 많이 있지만 사용한 입력력 변수와 데이터가 다르기 때문에 기존 연구와 정확하게 비교, 평가하기 어렵다. 주식시장 예측 연구는 대부분 종목별 주가 예측에 관한 것이고, 최근(5년 이내) 주가지수 연계 ETF 상품이 생기면서 지수 예측에 대한 연구가 시도되고 있다[1, 2, 8-9]. 이 중 ETF 거래를 목적으로 한 지수 예측은 다층신경회로망을 사용한 연구로[2, 9], 입력력 변수와 데이터의 시기는 다르지만 본 논문에서 구현한 MLPC와 예측 방법이 유사해서 비교 평가해 볼 수 있다. ETF 거래 손실을 줄이려면 모델의 정밀도가 높

아야 하는데, Table 2와 3을 보면 MLPC의 정밀도는 본 논문의 다른 모델 보다 뛰어나지 않다. 기존의 다층 신경 회로망 모델 보다 상승 예측에서는 SVC와 RFC가, 하락 예측에서는 SVC와 XGBC가 더 우수한 머신러닝 모델임을 볼 수 있다.

**Table 2. Accuracy and precision of up prediction**

Up prediction models	2019 ~ 2020	
	Accuracy (%)	Precision (%)
MLPC	74.0	61.1
SVC	72.3	68.4
RFC	74.1	68.0
LGBMC	73.5	40.0
XGBC	73.3	52.9

**Table 3. Accuracy and precision of down prediction**

Down prediction models	2019 ~ 2020	
	Accuracy (%)	Precision (%)
MLPC	79.0	40.5
SVC	79.4	50.0
RFC	76.0	35.5
LGBMC	77.7	14.3
XGBC	79.5	43.8

## 4.2 하이브리드 머신러닝 모델

최적의 매개변수 탐색을 통해 구현된 상승 예측 모델에서 가장 낮은 정밀도를 보인 LGBMC를 제외한 MLPC, SVC, RFC 및 XGBC로 하이브리드 상승 예측 모델을 구성하고, 하이브리드 하락 예측 모델은 43.8 % 이상의 정밀도를 보인 SVC와 XGBC만으로 구성한다. Table 4는 KOSPI가  $\alpha$  이상 상승하거나  $-\alpha$  이하로 하락한 실제 날의 수와 예측된 날의 수를 비교한다. 실제 상승과 하락한 일수를 보면 대략 2018년은 하락장, 2019년은 박스권이고 2020년은 상승장임을 추정할 수 있다. 예측 대상인 실제 날의 수가 전체 주식 개장일의 20 ~ 30 %로 작으므로 하이브리드 모델의  $\beta$ 를 1로 한다. 즉, 하나 이상의 모델에서 상승 또는 하락 예측이 있으면 ETF를 거래한다. Table 4에서 하락 예측 횟수가 상대적으로 적은 것은 하이브리드 하락 예측 모델에 포함된 개별 모델의 수가 적기 때문이다. Table 2, 3 및 5를 비교하면 상승 예측 모델보다 정밀도가 낮은 하락 예측 모델의 경우 하이브리드 모델의 정밀도가 단일 모델 보다 최소 14.3 %, 최대 20.5 % 개선되었다. 개별 모델에서 상승 또는 하락으로 예측된 날이 동일하지 않기에 이들의 예측 결과를 조합해 사용하면 예측 정밀도

가 개선됨을 볼 수 있다.

Table 4. No. of predicted days for trading ETF

Days		Verification		Test	
		2018	2019	2020	
Up	Actual	40	43	86	
	Predicted	20	8	57	
Down	Actual	54	36	65	
	Predicted	4	5	9	

Table 5. Accuracy and precision of hybrid prediction

Hybrid model	2019 ~ 2020	
	Accuracy (%)	Precision (%)
Up prediction	72.1	63.8
Down prediction	79.8	64.3

Table 6. Return expected by ETF trading

Expected return	Verification			Test		
	2018	2019	2020	2018	2019	2020
Up (%)	9.64	2.28	23.63	9.64	2.28	23.63
Down (%)	4.75	2.96	7.53	4.75	2.96	7.53
Total (%)	14.39	5.24	31.16	14.39	5.24	31.16

하이브리드 모델을 사용해 다음 날 KOSPI 종가의 방향을 예측하고 이를 토대로 KOSPI ETF 거래를 다음과 같이 시뮬레이션 한다. 다음 날에 대한 하이브리드 상승 및 하락 예측 모델의 출력이 각각 1과 0이면 다음 날 KOSPI 종가는 상승으로 예측되고 당일 증가 기준으로 KOSPI 연계 ETF를 매수한 후 다음 날 증가 기준으로 이를 매도한다. 다음 날에 대한 하이브리드 모델의 출력이 위와 반대로 0과 1이면 다음 날 KOSPI 종가는 하락으로 예측되고 당일 증가 기준으로 KOSPI 연계 인버스 ETF를 매수한 후 다음 날 증가 기준으로 이를 매도한다.

하이브리드 모델의 예측을 토대로 ETF를 거래할 경우 2018년, 2019년 및 2020년의 추정 수익률을 일별로 표시한 것이 Fig. 7, Fig. 8 및 Fig. 9이고 연도별 최종 수익률은 Table 6과 같다. ETF 거래 수수료는 0.1%로 계산한다. Fig. 7에서 KOSPI의 하락 추세에도 ETF 누적 거래 수익률이 증가함을 볼 수 있다. Fig. 8에서 KOSPI의 박스권 추세에서 상승과 하락 예측 횟수가 적어 ETF 누적 거래 수익률이 작았다. Fig. 9에서 KOSPI의 급락 후 대세 상승 추세에서, 갑작스런 KOSPI 급락으로 예측이 틀려 ETF 누적 거래 수익률이 -11.35%로 떨어졌다가 상승해서 31.16%가 됨을 볼 수 있다.

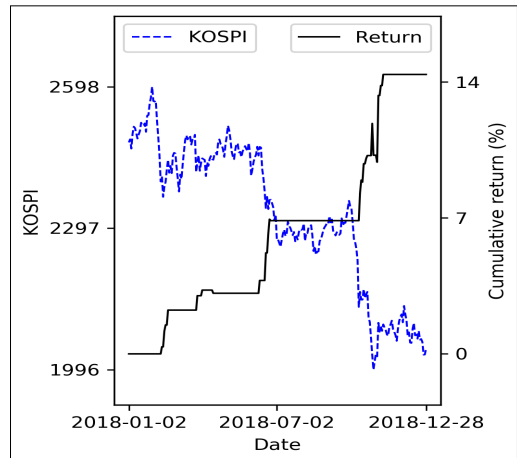


Fig. 7. Cumulative return from ETF trading in 2018

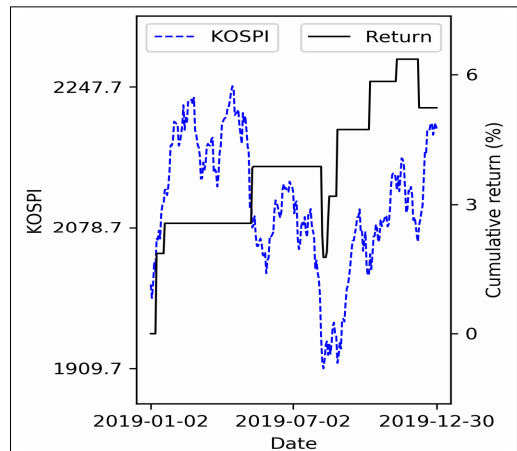


Fig. 8. Cumulative return from ETF trading in 2019

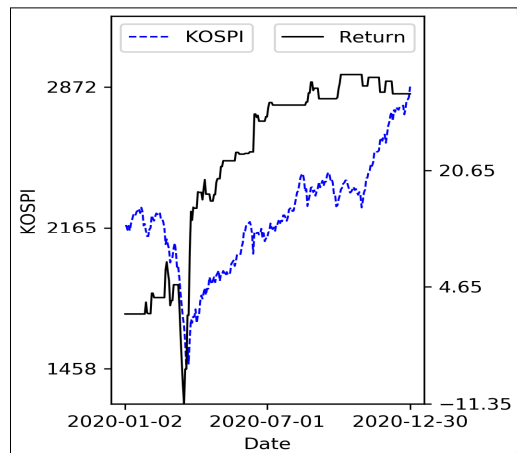


Fig. 9. Cumulative return from ETF trading in 2020

## 5. 결론

본 논문에서는 과거 8일 동안의 KOSPI 증가 데이터를 입력으로 다음 일의 KOSPI 증가 방향을 예측하는 머신러닝 모델을 제안한다. MLPC, SVC, RFC, LGBMC 및 XGBC를 사용해서 상승 및 하락 예측 모델을 각각 구현한다. 개별 예측 모델 보다 개선된 예측 정밀도를 얻기 위해 하이브리드 예측 모델을 제안한다. 하이브리드 상승 예측 모델은 MLPC, SVC, RFC, 및 XGBC를, 하이브리드 하락 예측 모델은 SVC와 XGBC를 조합해 구성한다. 하이브리드 모델에 대한 성능평가를 보면 개별 모델의 정확도와 정밀도가 모두 50 %를 넘는 경우, 이를 조합한 하이브리드 모델은 개별 모델의 평균적인 성능을 보이지만, 개별 모델의 정밀도가 50 % 보다 낮은 경우, 하이브리드 모델의 정밀도는 최소 14.3 % 개선됨을 볼 수 있다. 하이브리드 모델 예측을 토대로 ETF 거래를 시뮬레이션 한 결과 2018년 14.39 %, 2019년 5.24 %와 2020년 31.16%의 누적 수익률을 보인다. KOSPI가 박스권 추세인 2019년에 총 13회의 거래(상승 8회, 하락 5회)에서 거둔 5.24 %의 수익률은 만족스럽지 못하지만, KOSPI 기초지수의 일간 변동률에 2배로 연동해 움직이는 KOSPI 레버리지와 인버스×2를 거래한다면 수익률은 제시된 값의 1.5 ~ 2배가 될 것이다. 향후 하락 예측에 보다 적합한 머신러닝 알고리즘이 탐색되면 예측 정밀도와 ETF 거래 수익률을 향상시킬 수 있을 것으로 기대한다.

## REFERENCES

- [1] H. Amin, H. M. Moein & E. Morteza. (2016). Stock market index prediction using artificial neural network, *Journal of Economics, Finance and Administrative Science*. DOI : 10.1016/j.jefas.2016.07.002.
- [2] H. S. Hwang. (2019). ETF Trading Based on Daily KOSPI Forecasting Using Neural Networks, *Journal of the Korea Convergence Society*, 10(1), 7-12. DOI : 10.15207/JKCS.2019.10.4.001
- [3] L. D. Persio & O. Honcha. (2017). Recurrent Neural Networks Approach to The Financial Forecast of Google Assets, *Int. J. of Mathematics and Computers in Simulation*, 11, 7-13.
- [4] X. Ding, Y. Zhang, T. Liu & J. Duan. (2015). Deep Learning for Event-driven Stock Prediction, *Proc. of the 24th Int. Joint Conf. on Artificial Intelligence*, (pp. 2327-2333).
- [5] O. Ican & T. B. Celik. (2017). Stock Market Prediction Performance of Neural Networks: A Literature Review, *International Journal of Economics and Finance*, 9(11), 100-108.
- [6] M. Qiu & Y. Song. (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. *PloS one*, 11(5), e0155133. DOI : 10.1371/journal.pone.0155133
- [7] Y. Jiao & J. Jakubowicz. (2017). Predicting Stock Movement Direction with Machine Learning: An Extensive Study on S&P 500 Stocks. DOI: 10.1109/BigData.2017.8258518
- [8] J. K. S. Liew & B. Mayster. (2017). Forecasting ETFs with Machine Learning Algorithms. *The Journal of Alternative Investments*, 20(3), 58-78. DOI : 10.2139/ssrn.2899520.
- [9] H. S. Hwang. (2019). Predicting The Direction of The Daily KOSPI Movement Using Neural Networks For ETF Trades, *Journal of the Korea Convergence Society*, 10(4), 1-6. DOI : 10.15207/JKCS.2019.10.4.001
- [10] I. K. Nti, A. F. Adekoya & B. A. Weyori. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction, *Journal of Big Data*, 7(20), 1-40. DOI : 10.1186/s40537-020-00299-5
- [11] S. Mehta, P. Rana, S. Singh, A. Sharma & P. Agarwal. (2019). Ensemble Learning Approach for Enhanced Stock Prediction, *12th Int. Conf. on Contemporary Computing(IC3)*, (pp. 1-5). DOI : 10.1109/IC3.2019.8844891.
- [12] Z. Xiao & E. David. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms, *Financial Innovation*. 5, 5-24. DOI : 10.1186/s40854-019-0138-0.
- [13] G. Samuel, N. Philip & I. Dennis. (2019). On Stock Market Movement Prediction Via Stacking Ensemble Learning Method, *IEEE Conf. on Computational Intelligence for Financial Engineering & Economics*. (pp. 1-8). DOI : 10.1109/CIFEr.2019.8759062.
- [14] Y. Li & Y. Pan. (2020). A Novel Ensemble Deep Learning Model for Stock Prediction Based on Stock Prices and News. *arXiv:2007.12620*
- [15] B. Weng, L. Lu, X. Wang, F. M. Megahed & W. Martinez. (2018). Predicting Short-Term Stock Prices using Ensemble Methods and Online Data Sources, *Expert Systems with Applications*, 112, 258-273. DOI : 10.1016/j.eswa.2018.06.016
- [16] A. Singh. (2018). *A Comprehensive Guide to Ensemble Learning (with Python codes)*. <https://medium.com/analytics-vidhya/a-comprehensive-guide-to-ensemble-learning-with-pythhon-codes-5261650bb531>

- [17] K. H. Lee & G. S. Jo. (1999). Expert System for Predicting Stock Market Timing Using A Candlestick Chart. *Expert System With Applications*, 16, 357-364.

황 희 수(Heesoo Hwang)

[정회원]



- 1988년 2월 : 연세대학교 본대학원 전기과 (공학석사)
- 1993년 2월 : 연세대학교 본대학원 전기과 (공학박사)
- 1993년 4월 ~ 2001년 2월 : 한국고속철도기술개발사업단 선임 및 수석연구원
- 2001년 3월 ~ 현재 : 한라대학교 ICT공학부 교수
- 관심분야 : 머신러닝, 헬스케어 진단 및 예측, 데이터 과학
- E-Mail : hshwang@halla.ac.kr