Journal of the Korea Academia-Industrial cooperation Society
Vol. 24, No. 5 pp. 571-578, 2023

https://doi.org/10.5762/KAIS.2023.24.5.571
ISSN 1975-4701 / eISSN 2288-4688

랜덤 포레스트 모델을 활용한 채권형펀드 성과 예측 연구

문종윤, 김선웅, 최흥식*
국민대학교 비즈니스IT전문대학원

A Study on Predicting Performance of Fixed-Income Funds Using Random Forest Models

Jong Yoon Moon, Sun Woong Kim, Heung Sik Choi*
Graduate School of Business IT, Kookmin University

요 약 금융투자 상품을 선택할 때 투자자들은 최적의 포트폴리오를 구성하기 위해 어려움을 겪는다. 본 연구에서는 채권형펀드에 대한 투자자의 포트폴리오 선택을 돕기 위해 1개월 후의 상대적 성과를 예측하는 기계학습 모델을 제안하였다. 연구에서는 2017년 12월 1일부터 2023년 1월 31일까지의 채권형펀드 성과 자료와 국고채 금리, 원·달러 환율, 국내 성장률, 소비자물가지수 등 매크로 지표 데이터를 활용하여 랜덤 포레스트 모형을 구축하였고, 이를 통해 예측력을 검증하였다. 본 연구의 결과는 다음과 같다. 첫째, 제안 모델의 예측 결과는 90% 이상의 정확도와 정밀도를 보였다. 둘째, BorutaPy를 이용하여 특성을 선택했을 때 예측력이 상승하였다. 셋째, 교차검증을 통해 모델의 안정성을 확인할 수 있었다. 이전의 학계 연구들은 개별 금융상품의 가격과 수익률을 직접 예측하는 것이 대부분이었지만, 본 연구는 분류기를 활용하여 상대적으로 우수한 성과 그룹을 선택할 수 있도록 가이드를 제공한다. 이러한 점에서 기존 연구들과는 많은 차별성이 있다. 따라서, 본 연구는 투자자들이 채권형펀드에 대한 투자 결정에 간접적으로 도움을 줄 것으로 기대된다. 향후 연구에서는 딥러닝, 서포트벡터머신(Support Vector Machine) 등 다양한 기계학습 모형들과 비교하여 최적의 예측모델을 개발하고, 정확도를 높일 필요가 있다.

Abstract When choosing financial investment products, investors struggle to construct an optimal portfolio. In this study, we propose a machine learning model that predicts the relative performance of fixed income funds one month out in order to help investors make portfolio choices. To verify the predictive power, we build a random forest model using performance data on fixed income funds from December 1, 2017, to January 31, 2023, as well as macro indicator data such as government bond interest rates, the won-to-US-dollar exchange rate, the domestic growth rate, and the consumer price index. The results of this study are as follows. First, the prediction results show accuracy and precision of more than 90%. Second, prediction power increases when characteristics are selected using BorutaPy, and third, cross-validation confirms the stability of the model. While most previous academic studies directly predict the price and return on investment from individual financial instruments, this study utilizes a classifier to guide the selection of relatively good performance groups. In this respect, it differs from previous studies. Therefore, this study is expected to indirectly help investors make investment decisions about fixed income funds. For future research, it is necessary to develop an optimal prediction model and to improve accuracy by comparing various machine learning models, such as deep learning and the support vector machine.

Keywords : Machine Learning, Random Forest, Deep Learning, Support Vector Machine, Fixed Income Funds

*Corresponding Author : Heung Sik Choi(Kookmin Univ.)
email: hschoi@kookmin.ac.kr
Received April 10, 2023 Revised May 2, 2023
Accepted May 12, 2023 Published May 31, 2023

1. 서론

금융투자협회 일일펀드시장 동향 보고서에 따르면, 2012년 말 펀드시장의 총 금액은 약 317.3조원 이였고, 2022년 말에는 총 831.1조원으로 시장이 161.9% 성장하였다. Fig. 1은 주식형펀드와 채권형펀드의 금액을 보여주고 있는데, 각각 2012년 말 94.6조원, 46.9조원에서 2022년 말 91.4조원, 120조원으로 주식형펀드는 -3.4% 역성장을 하였고, 채권형펀드는 155.9% 성장을 하였다.

채권형펀드의 급성장은 한국, 미국, 유럽, 일본 등 각국 중앙은행의 저금리 정책 영향에 기인했지만, 2023년 3월 실리콘밸리은행(Silicon Valley Bank)의 파산과 크레디트스위스(Credit Suisse)의 몰락 등으로 인해 안정적인 투자처를 선호하는 자금 수요가 증가하면서, 채권형펀드는 지금도 여전히 고객들이 선호하는 금융상품으로 자리 잡고 있다.

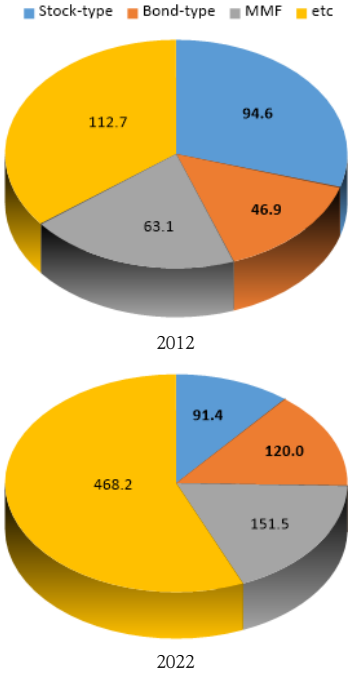


Fig. 1. Amount by fund type

Kim(2008)은 2001년부터 2006년까지 국내 채권형펀드의 알파 값 회귀분석을 통해 성과 지속성을 실증하였고[1], Huij and Derwall(2008)은 1990년부터 2003년까지 미국 채권형펀드의 상대적 성과에 대한 지속성을 연구하였다[2]. 또한, Clare et al. (2019)은 1998년 1월부터 2017년 2월까지 장기간에 걸친 채권형펀드의 포괄적인 데이터를 사용하여 미국 채권 뮤추얼 펀드 산업의 성과를 평가하면서, 주식형펀드에 비해 상대적으로 주목받지 못했던 채권형펀드에 대한 평가에 기여를 하였다[3].

Yong and Yang(2022)은 최근 채권형펀드에 대한 연구로, 국내 채권 보유내역 자료를 활용하여 성과를 분석하였다[4]. 또한, Otero-Gonzalez et al.(2022)은 글로벌 투자리서치 회사인 모닝스타의 스타(star) 등급이 유럽 채권 펀드시장에서 유용한지를 실증적으로 분석하였다[5].

주식형펀드에 관한 연구로는 Park and Chang (2001)이 국내 주식형펀드 시장의 성과를 평가하기 위한 연구를 수행하였고[6], Park(2007)은 주식형펀드 수익률에 영향을 미치는 결정요인들을 통계적으로 분석하였으며[7], Lee et al.(2014)은 국내 주식형펀드의 현금흐름과 거시경제변수 간의 관계를 분석하여, 현금흐름이 미래 경제 상황을 예측하는데 높은 설명력을 가지는 것을 확인하였다[8].

또한, Yim(2019)은 주식형펀드의 설정액과 주식시장 간의 영향을 실증적으로 조사하였고[9], Ha (2019)는 다요인 모형과 횡단면 분석을 활용하여 주식형 펀드의 시장초과 수익률을 연구하였으며[10], Wagner et al.(2022)은 주식형펀드의 계절적 특성을 분석하였다[11].

그동안의 국내외 학계에서는 주식형펀드에 대한 성과 연구들은 지속해서 이루어져 왔으나, 채권형펀드에 대한 연구들은 드물었다. Clare et al.(2021)은 본인의 연구논문에서 주식형펀드 보다 채권형펀드가 적은 이유에 관해서 설명하였다[12].

투자자에게 있어서 금융상품의 가격과 수익률을 정확하게 예측하고 분석하는 것은 투자의 핵심이며 가장 중요한 요소이다. 이에 관한 다양한 연구들이 진행됐었는데, Fama and French(1993)는 주식과 채권의 수익률에서 다섯 가지의 일반적인 위험 요소를 밝혔으며[13], Bhojraj and Sengupta(2003)는 채권의 수익률과 등급 간의 연관성을 연구하였고, 기업의 채무불이행 위험이 채권 수익률과 등급에 영향을 미칠 수 있음을 보여주었다[14].

Greenwood and Hanson(2013)은 채권 초과 수익률에 대한 발행기관 신용 요인의 유의 한 예측 가능성을 보여주며[15], Lin et al.(2014)은 1973년 1월부터 2010년 12월까지 회사채 초과 수익률의 예측 가능성을 조사하였고, 채권 유동성 요인과 신용 스프레드가 초과 수익률의 예측력을 검증하였다[16]. 또한, Devpura et al.(2021)은 시계열과 패널 예측 회귀 모델의 두 가지 방법을 사용하여 OECD 25개국의 채권 초과 수익률 예측 가능성을 검토하였으며 주식 수익률 예측 가능성보다 채

권 초과 수익률 예측 가능성에 관한 연구는 제한적이라고 주장하였다[17]. 그리고 Lee(2022)는 채권 수익률을 예측할 때 가장 중요한 경제적 불확실성이 무엇인지를 보여주었다[18].

본 연구는 채권형펀드와 펀드 수익률의 예측에 관한 연구로써 해외에서는 학술 연구들이 많이 이루어지고 있는 것과는 달리 국내는 문헌이 부족한 상황에서 진행한 논문으로서 의의가 있다고 볼 수 있다.

연구에 사용된 시계열 분석은 가장 전통적인 수익률 예측 방법 중 하나다. 시계열 분석은 과거 데이터로부터 미래의 값을 예측하는 분석 방법으로, 자기회귀(AR) 모형, 이동평균(MA) 모형, AR과 MA 모형을 혼합시켜 자기회귀이동평균(ARMA) 모형이 있다. Box and Jenkins (1976)는 비정상적(nonstationary) 시계열 자료에 대해 분석하는 진일보한 ARIMA 모형을 개발하였다[19].

최근 ARIMA 분석 방법에 관한 연구로, Ariyo et al.(2014)은 ARIMA 모델을 사용하여 주식의 가격을 예측하는 데 활용하였으며[20], Madhuri et al.(2020)은 Facebook 데이터 사이언스 팀에서 개발한 예언자(prophet)를 활용하여 ARIMA 시계열 분석에 기반을 둔 주식시장 예측을 수행하였다[21]. 또한, GARCH 분석 방법을 이용한 시계열 분석에 관한 최근 연구로, Suh(2016)는 GARCH 모델과 결합한 랜덤 포레스트 모형을 사용하여 환율을 예측하였고[22], Kim and Li(2022)는 GARCH 모형을 활용하여 중국 금 선물시장의 가격을 분석하고 예측하였다[23].

하지만, ARIMA와 GARCH 모델은 시계열 데이터의 단기 예측에는 잘 작동하지만, 장기적인 예측에는 성능이 떨어진다는 단점이 있고, 특히 다른 변수와의 인과 관계나 외부 요인에 대한 영향력 등을 분석하기 어렵다는 한계점이 있다.

현대 금융시장은 매우 복잡한 구조를 갖추고 있으며, 다양한 변수들에 의해 영향을 받고 있다. 학계에서는 이러한 구조를 파악하고 예측력을 높이기 위해 머신러닝 모형을 활용한 연구가 점점 더 활발해지고 있다. 이러한 기계학습 모형은 다양한 변수 간의 복잡한 패턴을 탐지하여 예측 정확도를 높일 수 있으므로 최근 금융시장에서는 매우 유용하게 활용되고 있는 방법론이다.

대표적인 기계학습 모형으로는 랜덤 포레스트 모델이 있는데 Ho(1995)가 처음으로 소개하였으며, 특성의 무작위 부분 공간에서 자란 트리의 컨센서스를 사용한다[24]. Kleinberg(1990, 1996)는 확률적 섭동과 평균화를 사용하여 과적합을 피하는 개념을 제안하였고[25,26], Amit and Geman(1997)은 이미지 분류 문제를 위해 이미지 특성에서 자란 임의화 된 트리를 사용하였다[27].

Breiman(1996)은 랜덤 포레스트의 전신인 배깅(bagging)을 소개하였고[28], Dietterich(2000)는 추가적인 임의화를 사용하여 배깅을 개선하였다[29]. Friedman and Hall(2007)은 부표본 추출이 배깅의 효과적인 대안임을 보여주었다[30].

Geurts et al.(2006)은 극단적으로 무작위한 트리의 랜덤 포레스트를 익스트림 랜덤 트리(extremely randomized trees) 또는 줄여서 엑스트라 트리(extra-trees)라 부르며, 일반적인 랜덤 포레스트보다 빠른 속도로 널리 활용되고 있다[31].

랜덤 포레스트의 최근 연구로는 Brown and Mues(2012)가 랜덤 포레스트는 빈도가 불균형한(imbalanced) 이진 분류(binary classification) 예측에서 우수한 성과를 보이는 것으로 보고하였다[32]. Uem and Kim(2022)은 의사결정 수 학습 알고리즘에 기반을 둔 랜덤 포레스트는 의사결정 나무의 앙상블(ensemble) 학습 기법으로 분석 데이터의 크기가 충분히 크지 않더라도 좋은 예측 성능을 보여주었다[33].

본 연구에서는 기계학습 모델을 활용하여 채권형펀드의 다음 달 수익률이 상위 그룹에 속할지, 하위 그룹에 속할지를 예측하고 분류하는 방법을 제안하였다. 이를 통해 수익률이 하위 50%에 속하는 펀드를 제외함으로써 최악의 선택을 회피할 수 있도록 경제적 가치를 제공한다.

Kang et al.(2019)은 펀드 투자와 관련된 의사결정을 지원하는 모델을 제시한 연구는 매우 부족하다고 언급하였다[34]. 이번 연구는 기존 학문에서 수익률을 직접 예측하는 것이 중심이었다면, 간접적으로 투자 결정을 돕는 차별성 있는 연구라고 볼 수 있다. 또한, 채권형펀드를 대상으로 기계학습 모델을 적용한 사례는 국내외적으로 연구가 부족한 실정이다.

2장에서는 본 연구의 주된 방법론이자 이진 분류 예측에서 우수한 성과를 보이는 랜덤 포레스트 모델로 적용될 자료 및 연구 모형을 소개하며, 그 실험 결과는 3장에서 분석한다. 4장에서는 본 연구에서 어떠한 학술적, 사회적, 기술적 가치를 얻을 수 있는지를 확인하고, 향후 연구 방향을 제시하고자 한다.

2. 자료 및 연구 모형

2.1 자료

본 연구에서는 2017년 12월 1일부터 2023년 1월 31일까지 한국펀드평가에서 일별로 제공한 143개 채권형 펀드의 가격(price)과 현금흐름(cashflow), 총보수(total cost)를 성과 데이터 자료로 사용하였고, 외생 변수 영향을 고려하기 위해서 연합인포맥스에서 제공한 한국 국고채 금리(3년, 10년), 원달러 환율, 한국 GDP 성장률, 한국 경기선행지수(OECD 발표 자료), 한국 소비자물가지수(CPI), 한국 실업률, 한국 GDP 성장률 전망치, 한국 취업자 수 증감 전망치, 한국 실업률 전망치, 한국 소비자물가지수 전망치를 매크로 데이터로 적용하였다. Table 1은 본 연구에서 사용한 자료(data)들을 보여준다.

관측 대상 펀드들을 예측하는 이진 분류(binary classification)를 위해 Eq. (1)과 같이 예측 대상 펀드의 1개월 수익률의 백분위 수치가 50 이상이면 0(상위권 수익률 펀드 그룹), 50 미만이면 1(하위권 수익률 펀드 그룹)로 라벨링 하였다.

$$Label\ for\ each\ fund \quad (1)$$
$$= \begin{cases} 0, \text{if } monthly\ percentile\ rank \geq 50, \\ 1, \text{if } monthly\ percentile\ rank < 50 \end{cases}$$

Table 1. Input Data

	Data	
Performance Data of Funds	Price	1 month Return
		3 month Return
		6 month Return
		12 month Return
		3 month Standard Deviation
		6 month Standard Deviation
		12 month Standard Deviation
		1 month Maximum Drawdown(MDD)
		3 month MDD
		6 month MDD
		12 month MDD
	Cash Flow	1 month CF
		3 month CF
		6 month CF
		12 month CF
	Total Cost	
Macro Data	3-year Treasury Rate	
	10-year Treasury Rate	
	KRW/USD Exchange Rate	
	GDP Growth Rate (YoY)	
	Leading Economic Index (OECD)	
	Consumer Price Index (CPI)	
	Unemployment Rate	
	GDP Growth Rate Forecast	
	Employment Change Forecast	
	Unemployment Rate Forecast	
	Consumer Price Index Forecast	

2.2 연구 모형

본 연구에서는 랜덤 포레스트 모형을 실험하기 위해 구글 코랩(google colab)을 사용하였다. 구글 코랩은 파이썬 스크립트를 작성하고 실행할 수 있는 편리한 도구이다. Fig. 2은 본 실험의 설계(design)를 보여주고 있다.

랜덤 포레스트 모형을 구현하기 위해 사이킷런(scikit-learn), 데이터 조작 및 분석을 위해 판다스(pandas), 대규모의 다차원 배열을 쉽게 처리할 수 있도록 넘파이(numpy) 등 오프소스 파이썬 프로그래밍 언어용 라이브러리를 설치하였다.

기계학습 실행에 앞서 전처리(pre-processing) 과정으로 데이터의 불균형(imbalance) 문제를 해결하기 위해서 모든 클래스의 개수를 맞춰 주었고, 데이터가 중복되고 불필요한 자료들은 삭제(drop)시켰다. 뿐만아니라, 과적합(overfitting)과 훈련시간(training time) 단축을 위한 과정도 거쳤다.

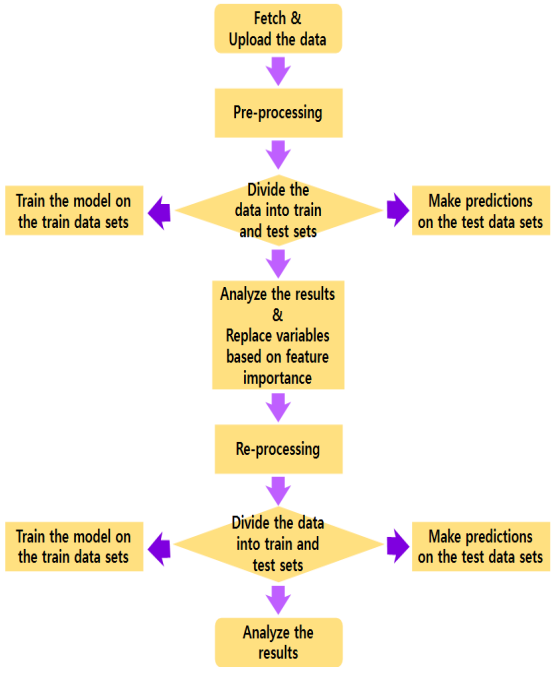


Fig. 2. Experimental design

모델링 전에는 훈련세트(train data set)와 테스트세트(test data set)를 나누어 주는데, 본 연구 모형에서는 훈련세트와 테스트세트를 7:3의 비율로 조절 하였다. 또

한 모델의 예측 능력을 높이기 위해서 하이퍼파라미터(hyperparameter)의 튜닝(tuning) 작업이 필요한데, Table 2은 본 모델에서 설정된 파라미터의 수치를 나타낸다.

Table 2. Set parameter values

Parameter	Value
n_estimators	100
max_depth	14
min_samples_split	2
min_samples_leaf	1
bootstrap	False
random_state	30

3. 실험 결과 분석

3.1 결정트리(decision tree) 분석

본 실험에서 결정 트리 개수는 100, 무작위로 선택할 난수의 개수는 30, 트리의 깊이는 14, 리프노드(leaf node)가 되기 위한 최소한의 샘플링 데이터 수는 1, 노드를 분할하기 위한 최소한의 샘플링 데이터 수는 2, 데이터 샘플링 중복 여부는 False를 가진다.

Fig. 3은 본 모형에서 실행한 결정 트리의 일부를 보여주고 있다. 상단의 트리에서는 측정할 자료(samples)가 총 9개가 있으며, 지니 불순도(gini impurity)는 0.346으로 나타낸다. 측정값이 표준화(standardization)된 실업률 지표 -0.703 보다 크다면 오른쪽 자식노드(child node)로 이동하며, 0(상위권 펀드 그룹)으로 예측한다. 이 경우 리프노드이므로 추가적인 검사를 하지 않는다.

만약, 상단 트리에서 측정값이 -0.703 보다 작거나 같다면 왼쪽 자식노드로 이동을 하고 표준화된 6개월 최대하락폭(MDD)을 측정하게 된다. 여기서 6개월 MDD가 0.12 보다 작거나 같다면 왼쪽 자식노드로 이동을 하고 예측치가 0, 샘플이 1개인 리프노드, 6개월 MDD가 0.12보다 크다면 오른쪽 자식노드로 이동을 하게 되고 예측치가 1(하위권 펀드 그룹)인 샘플이 7개 리프노드가 관측된다[35].

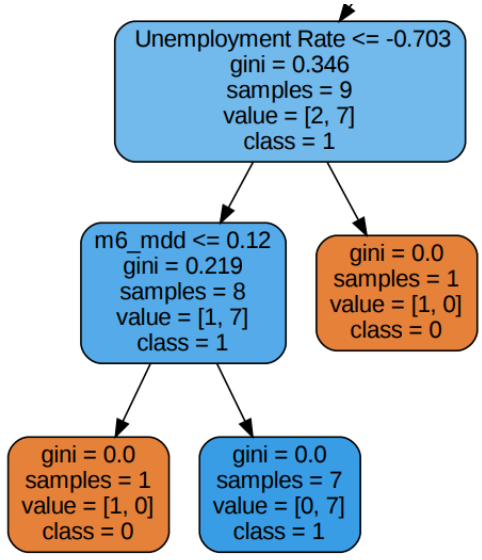


Fig. 3. Decision trees

3.2 특성 변수 중요도 분석

랜덤 포레스트는 특성의 상대적 중요도를 산출하기 쉬운 장점이 있다[36]. Fig. 4은 전체 27개 특성 변수 중요도(feature importance)의 표준편차를 나타내고 있으며, 오차막대(error bar) 차트로 표시하였다.

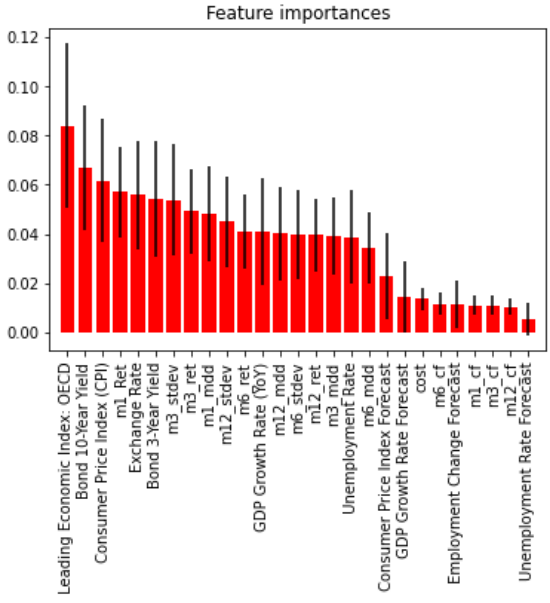


Fig. 4. Feature importances

본 모델에서 특성 중요도는 한국 경기선행지수(OECD 발표 자료), 한국 10년 국고채 금리, 한국 소비자물가지수(CPI), 1개월 수익률, 원달러 환율, 한국 3년 국고채 금리, 3개월 표준편차, 3개월 수익률, 1개월 최대 하락폭(MDD), 12개월 표준편차, 6개월 수익률, 한국 GDP 성장률 등의 순으로 나타났다.

3.3 모형 예측 결과

머신러닝 모형은 다양한 방법으로 예측 성능을 평가하는데, 본 연구에서는 분류 모델에 사용되는 평가지표를 사용하였다. Table 3은 실험 모형으로 테스트 데이터를 예측한 결과를 오차행렬(confusion matrix)로 나타낸 것이고, Table 4은 실험 모형의 예측성과를 표로 나타낸 것이다. 예측 평가지표는 Eq. (2)와 같이 계산된다[37].

Table 3. Confusion matrix of prediction

		Predict	
		0	1
Actual	0	756	76
	1	78	769

Table 4. Evaluation metrics of result

	Precision	Recall	F1-score
0	0.91	0.91	0.91
1	0.91	0.91	0.91
Accuracy			**0.9083**

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$\mathrm{Precision} = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = \frac{2\times \mathrm{Precision}\times Recall}{\mathrm{Precision}+Recall}$$

$$where$$
$$TP\ is\ true\ positive$$
$$FP\ is\ false\ positive$$
$$FN\ is\ false\ negative$$
$$TN\ is\ true\ negative$$

본 실험에서 정확도(accuracy)는 90.83%로 실제 데이터와 예측 데이터가 얼마나 일치하는지를 나타내는 수치로서, 이 연구의 모형은 뛰어난 예측 성과를 보여주고 있다. 정밀도(precision)와 재현율(recall)은 각각 91%로 데이터 분석에서 positive 데이터 세트의 성능을 평가하기 위한 지표이며, 분석 상황에 따라 이 중 하나를 선택하여 사용한다. F1 스코어는 정밀도와 재현율을 결합한 지표로, 두 값이 모두 균형을 이룰수록 높은 값을 가지며, 본 실험에서는 91%를 기록하였다.

3.4 모형 성능 향상(improvement)

파이썬에서 BorutaPy 패키지를 설치하면 랜덤 포레스트 모형에서 특성(feature) 변수를 중요도에 따라 줄여서 모형의 정확성을 높일 수가 있다. 본 연구에서도 BorutaPy 패키지를 활용하여 총 27개의 변수 중에서 4개를 줄여서 23개의 변수로 재설정 하였다. 또한 교차검증(cross validation)을 통하여 과적합(overfitting)과 과소적합(underfitting)을 탐지하여 더욱 일반화된 모델을 생성하였다. Table 5은 최종적으로 모형 성능을 향상한 후, 추가 시뮬레이션을 통한 평균 결과값인데, 정확도가 약 91.89%를 보여준다. 이는 최초 정확도 90.83% 보다 개선됨을 확인할 수 있다.

Table 5. Average value of simulations

	Precision	Recall	F1-score
0	0.918	0.920	0.920
1	0.920	0.918	0.920
Accuracy			**0.91886**

4. 결론

본 연구는 아래와 같은 학술적, 사회적, 기술적 가치를 지니고 있다고 판단된다. 첫째, 금융투자 상품을 직접 예측하는 기존 연구들과는 차별화된 학문적 가치를 제공하며, 둘째, 투자자들에게 보다 안전한 투자 결정을 도모함으로써, 투자 실패를 줄이고, 경제적 이익을 상승시키는 금융안정성이 향상되는 사회적 가치를 가진다. 셋째, 본 연구에서 도출된 모델로 인해 다양한 자산 분야에서의 머신러닝 기반 예측 및 분석 기술 확산이 기대되며, 금융 분야의 예측 기술 발전에 이바지할 수 있다.

끝으로 향후 연구 방향은 다음과 같다. 첫째, 다양한 금융상품에 대한 분석 확장이 필요하다. 본 연구에서는 채권형펀드만을 측정 대상으로 하였으나, 주식형펀드, 부동산펀드, 혼합형펀드 등 다른 자산도 본 모델을 적용할 경우 새로운 학술적인 성과를 보일 수 있을 것으로 기

대된다. 또한, 분류기를 적용하여 높은 성과를 보일 수 있는 다른 금융자산을 접목하는 연구도 필요하다.

둘째, SVM, Deep learning 등의 다른 머신러닝 모형을 이용하여 분석해 보는 것이 필요하다. 또한, 각 모형의 장단점을 분석하고, 본 모델과 비교하여 더욱 정확하고 효과적인 분석 방법을 모색할 필요가 있다.

셋째, 본 연구에서는 펀드 성과지표 16개와 매크로 지표 11개를 활용(총 27개 특성 변수)하였지만, 예측 자산과 상관관계가 높은 변수들을 추가로 발굴하여 연구에 접목한다면 더욱 발전된 결과를 얻을 수 있을 것으로 기대된다.

References

[1] H. B. Kim, "Performance persistence of fixed income funds", *Korean Journal of Business Administration*, Vol.21, No.2, pp.567-585, 2008.

[2] J. Huij, J. Derwall, "Hot hands in bond funds", *Journal of Banking & Finance*, Vol.32, pp.559-572, Apr. 2008. DOI: https://doi.org/10.1016/j.jbankfin.2007.04.023

[3] A. Clare, N. O'Sullivan, M. Sherman, S. Zhu, "The performance of US bond mutual funds", *International Review of Financial Analysis*, Vol.61, pp.1-8, Jan. 2019. DOI: https://doi.org/10.1016/j.irfa.2018.12.001

[4] H. S. Yong, C. W. Yang, "Performance of korean bond mutual funds: analysis using fund holding data", *Korean Financial Management Association*, Vol.39, No.6, pp.105-129, 2022.
DOI: https://doi.org/10.22510/kjofm.2022.39.6.004

[5] L. Otero-Gonzalez, P. Leite, P. Duran-Santomil, R. Domingues, "Morningstar star ratings and the performance, risk and flows of european bond mutual funds", *International Review of Economics & Finance*, Vol.82, pp.479-496, Nov. 2022.
DOI: https://doi.org/10.1016/j.iref.2022.07.003

[6] Y. K. Park, U. Chang, "The performance measurement and attribute analysis on equity style funds", *The Journal of Financial Studies*, Vol.29, pp.117-143, 2001.

[7] B. J. Park, "The determination factors of mutual fund return", *The korean journal of financial management*, Vol.24, pp.85-107, Mar. 2007.

[8] H. D. Lee, S. B. Kim, E. J. Yoo, "Relationship between macroeconomic variables and fund flow in the korean equity funds", K*orean association of industrial business administrarion*, Vol.29, pp.175-197, May. 2014.

[9] B. J. Yim, "A study on the mutual influence of the stock fund provisions and stock market in korea", *The Journal of Business Education*, Vol.33, No.5, pp.145-162, 2019.
DOI: https://doi.org/10.34274/krabe.2019.33.5.007

[10] Y. J. Ha, "Market returns and equity fund performance", *The Korean Finance Association*, Vol.32, No.1, pp.55-92, 2019.
DOI: https://doi.org/10.37197/ARFR.2019.32.1.3

[11] M. Wagner, J. B. T. Lee, D. Margaritis, "Mutual fund flows and seasonalities in stock returns", *Journal of Banking & Finance*, Vol.144, Nov. 2022.
DOI: https://doi.org/10.1016/j.jbankfin.2022.106623

[12] A. Clare, K. Cuthberson, D. Nitzsche, N. O'Sullivan, "How skilful are US fixed-income fund managers?", *International Review of Financial Analysis*, Vol.74, Mar. 2021.
DOI: https://doi.org/10.1016/j.irfa.2021.101673

[13] E. F. Fama, K. R. French, "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, Vol.33, pp.3-56, Feb. 1993.
DOI: https://doi.org/10.1016/0304-405X(93)90023-5

[14] S. Bhojraj, P. Sengupta, "Effect of corporate governance on bond ratings and yields", *The Journal of Business*, Vol.76, No.3, pp.455-475, Jul. 2003.
DOI: https://doi.org/10.1086/344114

[15] R. Greenwood, S. G. Hanson, "Issuer quality and corporate bond returns", *The Review of Financial Studies*, Vol.26, pp.1483-1525, Jun. 2013.
DOI: https://doi.org/10.1093/rfs/hht016

[16] H. Lin, J. Wang, C. Wu, "Predictions of corporate bond excess returns", *Journal of Financial Markets*, Vol.21, pp.123-152, Nov. 2014.
DOI: https://doi.org/10.1016/j.finmar.2014.08.003

[17] N. Devpura, P. K. Narayan, S. S. Sharma, "Bond return predictability: evidence from 25 OECD countries", *Journal of International Financial Markets, Institutions and Money*, Vol.75, Nov. 2021.
DOI: https://doi.org/10.1016/j.intfin.2021.101301

[18] K. Lee, "Which uncertainty measures matter for the cross-section of corporate bond returns? Evidence from the U.S. during 1973-2020", *Finance Research Letters*, Vol.48, Aug. 2022.
DOI: https://doi.org/10.1016/j.frl.2022.102913

[19] G. E. P. Box, G. M. Jenkins, "Time series analysis: forecasting and control", *2nd ed., San Francisco: Holden-Day*, 1976.

[20] A. A. Ariyo, A. O. Adewumi, C. K. Ayo, "Stock price prediction using the ARIMA model", *UKSim-AMSS 16th International Conference on Modelling and Simulation*, pp.106-112, 2014.
DOI: https://doi.org/10.1109/UKSim.2014.67

[21] C. Madhuri, M. Chinta, P. Kumar, "Stock market prediction for time-series forecasting using prophet upon ARIMA", *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, 2020.
DOI: https://doi.org/10.1109/ICSSS49621.2020.9202042

[22] J. D. Suh, "Foreign exchange rate forecasting using the GARCH extended random forest model", *Korean Industrial Economic Association*, Vol.29, pp.1607-1628,

Oct. 2016.

[23] S. T. Kim, M. H. Li, "Volatility analysis and prediction based on ARMA-GARCH-type models: evidence from the chinese gold futures market", *Korea Trade Research Association*, Vol.47, pp.212-235, Jun. 2022.

[24] T. K. Ho, "Random decision forests", *Proc. Third International Conference on Document Analysis and Recognition, IEEE Computer Society Press, New York*, Vol. 1, pp.278-282, 1995.
DOI: https://doi.org/10.1109/ICDAR.1995.598994

[25] E. M. Kleinberg, "Stochastic discrimination", *Annals of Mathematical Artificial Intelligence*, Vol.1, pp.207-239, 1990.
DOI: https://doi.org/10.1007/BF01531079

[26] E. M. Kleinberg, "An overtraining-resistant stochastic modeling method for pattern recognition", *Annals of Statistics*, Vol.24, pp.2319-2349, 1996.
DOI: https://doi.org/10.1214/aos/1032181157

[27] Y. Amit, D. Geman, "Shape quantization and recognition with randomized trees", *Neural Computation*, Vol.9, pp.1545-1588, Oct. 1997.
DOI: https://doi.org/10.1162/neco.1997.9.7.1545

[28] L. Breiman, "Bagging predictors", *Machine Learning*, Vol.24, pp.123-140, Aug. 1996.
DOI: https://doi.org/10.1007/BF00058655

[29] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization", *Machine Learning*, Vol.40, pp.139-157, Aug. 2000.
DOI: https://doi.org/10.1023/A:1007607513941

[30] J. H. Friedman, P. Hall, "On bagging and nonlinear estimation", *Journal of Statistical Planning and Inference*, Vol.137, pp.669-683, Mar. 2007.
DOI: https://doi.org/10.1016/j.jspi.2006.06.002

[31] P. Geurts, D. Ernst, L. Wehenkel, "Extremely randomized trees", *Machine Learning*, Vol.63, pp.3-42, Mar. 2006.
DOI: https://doi.org/10.1007/s10994-006-6226-1

[32] I. Brown, C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets", *Expert Systems with Applications*, Vol.39, No. 3, pp.3446-3453, Feb. 2012.
DOI: https://doi.org/10.1016/j.eswa.2011.09.033

[33] D. Y. Uem, S. Y. Kim, "A study on the analysis of factors for the golden glove award by using machine learning", *The Journal of the Korea Contents Association*, Vol.22, pp.48-56, 2022.
DOI: https://doi.org/10.5392/JKCA.2022.22.05.048

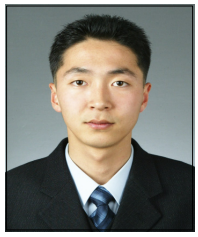
[34] S. S. Kang, H. J. Lee, K. J. Oh, "Using genetic algorithm to build an investment optimization model based on fund clustering", *Journal of the Korean Data And Information Science Sociaty*, Vol.30, No.2, pp.285-297, 2019.
DOI: https://doi.org/10.7465/jkdi.2019.30.2.285

[35] A. Geron, "Hands-on machine learning with scikit-learn & tensorflow", *Hanbit Media Inc.*, pp.225-239, 2018.

[36] S. Jansen, "Machine learning for algorithmic trading", *Seoul : Acon*, pp.256, 2021.

[37] A. C. Muller, S. Guido, "Introduction to machine learning with python", *Hanbit Media Inc.*, pp.376-406, 2017.

문 종 윤(Jong Yoon Moon) [정회원]



- 2002년 2월 : 경상대학교 전자공학과 (공학사)
- 2007년 3월 : University of York, Mathematical Finance (MSc)
- 2016년 3월 ~ 2018 8월 : 국민대학교 비즈니스IT전문대학원 박사과정 수료
- 2006년 12월 ~ 현재 : DGB금융그룹 부부장

〈관심분야〉
트레이딩시스템, 금융공학, 채권운용

김 선 웅(Sun Woong Kim) [정회원]

- 1981년 2월 : 서울대학교 경영학과 (경영학사)
- 1983년 2월 : KAIST 경영과학과 (공학석사)
- 1988년 2월 : KAIST 경영과학과 (공학박사)
- 2009년 3월 ~ 현재 : 국민대학교 비즈니스IT전문대학원 교수

〈관심분야〉
트레이딩시스템, 투자위험관리

최 흥 식(Heung Sik Choi) [정회원]

- 1983년 2월 : 한양대학교 산업공학과 (공학사)
- 1985년 2월 : KAIST 경영과학과 (공학석사)
- 1995년 2월 : 로체스터대 컴퓨터정보시스템 (경영학박사)
- 1995년 3월 ~ 현재 : 국민대학교 비즈니스IT전문대학원 교수

〈관심분야〉
트레이딩시스템, 증권투자연구