

金融素养与投资决策——基于机器学习的再研究

敖兴坡¹ 何琳²

1 杭州电子科技大学 经济学院

2 中南财经政法大学 工商管理学院

DOI:10.32629/ej.v9i1.3295

[摘要] 民间投资是金融市场的重要组成部分。本文采用机器学习的集成算法模型,深入考察金融素养对投资者投资决策的预测能力。结果表明,投资者的金融素养对投资决策有着较为显著的预测能力。实证分析表明,金融素养与是否投资无因果关系,但对投资盈亏有显著正向影响。异质性分析表明,高金融素养对是否投资及投资盈亏均起正向促进作用。但低金融素养则与二者呈负相关。基于此,本文对我国投资者投资行为提出相关建议,以期为健全投资风险监测预警、完善散户投资监管机制提供参考。

[关键词] 金融素养; 投资决策; 机器学习

中图分类号: F283 **文献标识码:** A

Financial Literacy and Investment Decisions: A Machine Learning Reexamination

Xingpo Ao¹ Lin He²

1 School of Economics, Hangzhou Dianzi University

2 School of Business Administration, Zhongnan University of Economics and Law

[Abstract] Private investment plays a crucial role in financial markets. This paper uses a machine learning ensemble model to examine how well financial literacy predicts investors' decisions. Results show that financial literacy significantly predicts investment behavior. Empirical analysis finds no causal link between financial literacy and the decision to invest, but a positive effect on investment returns. Heterogeneity analysis indicates that high financial literacy promotes both investment participation and performance, while low literacy correlates negatively with both. Based on this, the paper offers recommendations for Chinese investors, aiming to help improve investment risk monitoring and regulation of retail investors.

[Key words] Financial Literacy; Investment Decision-making; Machine Learning

引言

民间投资是金融市场发展的重要环节。国家发展改革委2023年7月发文强调要促进民间投资,保持其占固定资产投资的合理比重。但《2022个人投资者投资行为分析报告》显示,部分投资者亏损主要源于系统性理财知识的缺乏和投资心态的激进,且投龄越长,投资者越重视系统投资理财知识,“知识改变‘投运’”渐成共识,金融素养提升愈发必要。

金融素养重要性凸显有两方面原因:一是居民收入增长使更多人参与金融市场,而金融产品和服务数量增多、日趋复杂,要求投资者具备相应认知与实践能以做出审慎决策。二是居民金融素养水平关系到普惠金融发展,贫困地区金融消费者常因不熟悉金融体系、缺乏基础知识与风险意识,导致其主动获取金融服务的意愿较低。

在此背景下,结合金融素养指标与前沿机器学习算法,优化投资者投资决策预测模型,对其投资决策进行有效预警,具有现

实意义。这既能辅助投资者调整策略、识别自身风险薄弱环节,也可为完善金融投资风险处置机制提供参考。

1 文献综述

梳理文献发现,关于金融素养定义和概念的研究较为丰富。有学者认为金融素养是综合性概念,Lusardi et al. (2014)^[1]指出其涵盖金融知识及处理经济信息、做出明智决策的能力。更有学者结合行为金融学、心理学成果,从人类认知能力发展角度解读金融素养,探索金融教育的必要性^[2]。同时,部分文献聚焦金融素养对投资的影响,指出金融素养提升有助于提高青年人投资理财意愿和收益率。盛智明和蔡婷婷(2021)^[3]基于相关数据发现,金融从业经历通过金融素养影响家庭风险投资行为。国外学者认为,高金融素养投资者在金融市场参与度更高,能选择更合适的金融资产^[4]。可见,金融素养提升对投资者参与投资和风险管理有正向作用。

此外,投资预测识别方法也在不断优化。传统计量经济学模

型要求金融时间序列满足一定假设条件以保证结果可靠。但金融市场动态、复杂且非线性，机器学习与深度学习模型没有严格假设条件，处理非线性和非平稳特征问题能力更强。Phua等(2003)^[5]通过构建神经网络模型，以超60%的精度预测五大股指趋势；Tsai等(2009)^[6]用集成算法预测股票价格，F-score得分比单一算法高11%。鉴于此，本文采用机器学习预测投资者投资，挖掘影响投资者的重要因素。

综上所述，多数投资预测基于金融市场独有特征，鲜有学者从投资者自身角度剖析金融素养与投资决策的关系。因此，本文引入新机器学习算法识别投资决策，采用集成模型，考察金融素养对投资预测的预测能力。通过构建因果森林算法识别框架与probit模型，探究金融素养与投资决策的内在联系。

2 数据与模型设定

2.1 数据来源

本文的数据来源于2019年的“中国家庭金融调查”(CHFS)。调查范围覆盖全国29个省(自治区、直辖市)，170个城市，345个区县，1360个村(居)委会，样本规模达34643户。

考虑到预测变量的特殊性，本文对样本进行问卷筛选，同时剔除了关键信息缺失、信息不可用的个体样本，最后剩余1112条有效样本。

2.2 变量说明

2.2.1 预测变量。(1)投资盈亏：由于CHFS2019年调查问卷中“去年，您家从这些非人民币资产买卖中实际产生的盈亏情况？”属于四分类变量，本文考虑到机器学习预测准确性的问题，剔除了“持平”与“没有买卖”这两个回答的样本，最终设定“投资盈亏”为二分类定类变量，定义盈利为“1”，亏损为“0”。(2)是否投资：根据问题“目前，您家是否拥有理财产品？”，回答为“1”，否为“0”。

表1 变量说明

变量属性	指标	变量说明	赋值
核心解释变量 (金融素养)	金融经验	目前，您家是否有股票账户？	是=1, 否=0
		目前，您家是否持有基金？	是=1, 否=0
	金融信息	您平时对经济、金融方面的信息关注吗？	非常关注=5, 很关注=4, 一般关注=3, 很少关注=2, 从不关注=1
		您家是否有理财顾问或者投资顾问？	是=1, 否=0
		2018年9月28日后，多数银行理财产品的投资门槛从5万下降到1万，您知道这个信息吗？	是=1, 否=0
	金融知识	假设银行的年利率是4%，如果把100元钱存一年定期，一年后获得的本金和利息？	回答正确=1, 回答错误=0
		假设银行年利率是5%，通货膨胀率为8%，把100元钱存银行一年之后能够买到的东西？	回答正确=1, 回答错误=0
		您对股票、债券、基金的整体了解程度如何？	非常了解=5, 比较了解=4, 一般了解=3, 不太了解=2, 完全不了解=1

2.2.2 核心解释变量。金融素养：本文参考姜楠等(2023)^[7]的研究，将金融素养界定为金融经验、金融知识、金融信息三个

维度，一共八题来构建金融素养指标。具体说明见表1。

2.2.3 控制变量。影响个体投资决策的因素有很多，为了进一步考察户主个体特征，本文选取年龄、性别、受教育程度、身体状况、婚姻、政治面貌、风险态度作为控制变量。

2.3 模型设定和评估指标

本文采用由前馈神经网络、支持向量机、随机森林、梯度提升树、决策树五个学习器组成的机器学习集成算法进行预测。

在模型评估中，主要采用混淆矩阵中的准确率(Accuracy)、精准度(Precision)、召回率(Recall)、F1-score(F1)以及roc_auc_score(AUC)对模型的预测性能进行评估。准确率(Accuracy)：分类模型所有判断正确的结果占总观测值的比重。精准度(Precision)：模型预测是positive的所有结果中，模型预测对的比重。召回率(Recall)：真实值是positive的所有结果中，模型预测对的比重。F1-score(F1)：指标综合了Precision与Recall的产出的结果，取值越大表示模型输出越好。Roc_auc_score(AUC)：是ROC曲线下的面积，实际意义为从数据集中任选一个正样本和一个负样本，模型打分时将正例分数排在反例前面的概率。

3 集成算法的预测与probit模型回归

3.1 基于集成算法的目标变量预测结果

本文将1112份数据样本随机分为两份：一份80%的数据作为训练集，用来构建模型，另一份20%作为测试集用于模型的测试和评估，模型的评估结果显示在预测投资盈亏上，各学习器准确率较高，前馈神经网络、随机森林、决策树、梯度提升树准确率超0.8，集成模型准确率达0.941。在预测是否投资时，各个学习器的准确率大约在0.6，在最终的集成模型中将准确率提高到了0.707。但决策树和梯度提升树Recall和F1得分低，导致集成模型在这两个得分上的预测结果不尽如人意，其可能的原因是特征选择问题，由于决策树和梯度提升树是基于特征的划分来做决策的，如果选择的特征不合适或者特征的重要性评估不准确，可能导致模型在某些情况下无法捕捉到重要的特征信息，从而影响召回率和F1得分。

3.2 基于因果森林算法的识别框架与probit模型

本文使用“R-learner”函数，将因果树生成一片的广义随机森林，获得权重。

$$\tau(x) = E[Y_i(1) - Y_i(0) | X_i = x] \quad (1)$$

$$\hat{\tau} = \frac{\sum_{i=1}^n \alpha_i(x)[Y_i - \hat{m}^{(-i)}(X_i)][Z_i - \hat{e}^{(-i)}(X_i)]}{\sum_{i=1}^n \alpha_i(x)[Z_i - \hat{e}^{(-i)}(X_i)]^2} \quad (2)$$

其中， $\alpha_i(x)$ 是由机器学习驱动形成的核(kernel)， Z_i 为处理变量。用 $e(x) = P[Z_i | X_i = x]$ 表示倾向值得分，符号 $-i$ 表示“袋外(out-of-bag)”预测，即 Y_i 不用于计算 $\hat{m}^{(-i)}(X_i)$ ，金

融素养影响效果的期望值 $m(x) = E[Y_i | X_i = x]$ 。 $e(x)$ 是倾向值得分, 满足 $0 \sim 1$ 均匀分布。 $\hat{m}(\cdot)$ 和 $\hat{e}(\cdot)$ 是广义随机森林的估计值, 并在两个森林中进行“袋外”预测。

同时, 将因子分析后的金融素养进行四分位分组, 将前25%的金融素养因子得分标识为处理组, 取值为1, 后25%的金融素养因子得分标识为控制组, 取值为0。由于被解释变量为二值变量, 且不存在数据截断现象, 因此本文采用probit模型进行回归分析, 模型具体形式如下:

$$y^* = \partial_0 + \partial_1 fin + \partial_2 control + \varepsilon \quad (3)$$

$$y^* = \partial_0 + \partial_1 fin + \partial_2 risk + \partial_3 risk * fin + \partial_4 control + \varepsilon \quad (4)$$

其中, y^* 为被解释变量(预测变量), 分别代表投资盈亏以及是否投资, fin 为金融素养, 是核心解释变量, $risk$ 表示风险态度, $risk * fin$ 表示风险态度与金融素养的交乘项, $control$ 代表模型中的控制变量, 表示随机干扰项。

3.3 回归结果与分析

金融素养与投资盈亏和是否投资的probit回归分析结果显示, 金融素养对投资盈亏有显著的正向影响, 提升金融素养可以提高投资盈利的可能性。但金融素养对是否投资的影响不显著, 其可能的原因是除了金融素养外, 还存在许多其他决定投资行为的因素, 如个人收入水平、风险偏好、资金可获得性等。此外, 不同个体对投资的偏好因素存在差异, 一些人可能金融素养较低但偏好风险, 另一些人可能金融素养较高但本身不愿进行风险投资, 因此受访者做出是否投资的决策可能与自身拥有的金融素养水平无关。

因果森林算法的结果显示提高1%的金融素养能力, 能提高投资盈利概率14.3%, 而probit模型下结果显示提高1%的金融素养能力, 能提高投资盈利概率13.6%。两者结论再次验证了金融素养与投资盈亏成正相关关系。因果森林算法下的金融素养与是否投资的因果关系不显著, 同样印证了probit回归结果。

3.4 金融素养异质性分析

为了探究金融素养程度不同的群体, 其自身的因素对投资决策的预测能力, 本文将金融素养因子得分前40%的样本作为高金融素养群体, 将后60%的样本作为低金融素养群体, 进行异质性检验, 结果表明, 高金融素养对是否投资以及投资盈亏起到正向促进作用。但低金融素养与两者呈负相关, 究其原因低金融素养者在面对投资市场时会产生对未知领域的抵触, 由于不了解投资市场, 不知道如何选择投资产品, 或者缺乏获取相关信息的途径。这可能使他们在投资决策上感到不安, 特别是该领域与自身的利益息息相关。

4 结论与启示

本文从投资者的金融素养这一研究角度出发, 结合前沿的机器学习方法对投资者投资决策进行有效预测。基于前馈神经网络、支持向量机、随机森林、梯度提升树、决策树的集成算法模型的分析结果显示, 投资者金融素养与自身特征对投资决策有着较为显著的预测能力。probit回归结果表明金融素养与是否投资并无因果关系, 而金融素养对投资盈亏有显著的正向影响, 提升金融素养可以提高投资盈利的可能性。异质性分析表明, 高金融素养对是否投资以及投资盈亏起到正向促进作用, 提升投资者自身的金融素养水平, 能够促进投资以及增加盈利的概率。但低金融素养与两者呈负相关。

对于投资者而言, 要充分认识到金融素养在投资中的关键作用。投资者需主动加强金融知识学习, 提升自身金融素养, 科学地评估投资风险与收益, 做出合理投资决策。金融机构应发挥专业优势, 开展多样化的金融教育活动, 帮助投资者提升金融素养。同时, 依据投资者金融素养和特征, 提供个性化投资产品与服务。监管部门要重视金融素养普及工作, 将其纳入金融监管体系。制定相关政策鼓励金融机构开展投资者教育, 加强对低金融素养群体的保护, 规范金融市场秩序, 营造良好投资环境, 促进金融市场健康稳定发展, 实现投资者与市场的共赢。

【参考文献】

- [1]Lusardi, Annamaria and Mitchell, Olivia. "The Economic Importance of Financial Literacy: Theory and Evidence," [J]. Journal of Economic Literature, 2014, 52(1), pp.5-44.
- [2]刘国强. 我国消费者金融素养现状研究——基于2017年消费者金融素养问卷调查[J]. 金融研究, 2018, (03): 1-20.
- [3]盛智明, 蔡婷婷. 金融从业经历、金融素养与家庭风险投资——基于“中国家庭金融调查(CHFS)”2017数据的分析[J]. 东南大学学报(哲学社会科学版), 2021, 23(05): 79-88+151.
- [4]Hung A., Parker A M., Yoong J., "Defining and Measuring Financial Literacy" [J]. RAND Working Paper Series. 2009, WR-708.
- [5]Phua P K H, Zhu X, Koh C H. Forecasting Stock Index Increments Using Neural Networks with Trust Region Methods[J]. IEEE, 2003, 1: 260-265.
- [6]Tsai C F, Wang S P. Stock Price Forecasting By Hybrid Machine Learning Techniques[J]. Lecture Notes in Engineering & Computer Science, 2009, 755(1): 60.
- [7]姜楠, 张帅华. 金融素养对家庭收入影响研究——来自CHFS数据的实证分析[J]. 价格理论与实践, 2023, (02): 143-146+203.

作者简介:

教兴坡(1997--), 男, 汉族, 河南信阳人, 杭州电子科技大学经济学院研究生, 硕士, 研究方向: 数字金融, 金融投资。

何琳(2001--), 女, 汉族, 安徽池州人, 中南财经政法大学工商管理学院研究生, 硕士, 研究方向: 跨国金融投资。