

# 基于RF-Adaboost的企业财务困境预警研究

方逸雯\*, 刘媛华

上海理工大学管理学院, 上海

收稿日期: 2023年1月21日; 录用日期: 2023年2月16日; 发布日期: 2023年2月23日

## 摘要

针对390家上市企业样本, 本文首先从获利能力、现金流量、营运能力、发展能力、偿债能力五个维度选取了13个初始财务指标, 根据随机森林特征重要性筛选出了6个贡献度最大的最终财务指标。其次, 本文建立了基于随机森林特征降维和Adaboost分类预测的RF-Adaboost模型, 根据企业T-2年的财务指标预测其在T年是否会被特殊处理。实证结果表明RF-Adaboost模型在测试集上的分类正确率和召回率都达到80%以上。最后, 为了验证RF-Adaboost模型的效果, 本文还使用了Adaboost模型、LSTM神经网络、RBF-SVM、Linear-SVM、基于核密度估计的朴素贝叶斯模型进行实验, 研究结果表明RF-Adaboost在所有模型中表现最好, 说明了特征降维的有效性和集成算法的优越性。

## 关键词

财务困境预警, 随机森林, Adaboost

# Research on Financial Distress Warning of Enterprises Based on RF-Adaboost

Yiwen Fang\*, Yuanhua Liu

Business School of University of Shanghai for Science and Technology, Shanghai

Received: Jan. 21<sup>st</sup>, 2023; accepted: Feb. 16<sup>th</sup>, 2023; published: Feb. 23<sup>rd</sup>, 2023

## Abstract

Aiming at the samples of 390 listed enterprises, this paper firstly selects 13 initial financial indicators from the 5 dimensions of profitability, cash flow, operating capacity, development capacity and debt paying capacity, then selects 6 final financial indicators with the greatest contribution based on the random forest' feature importance. Secondly, this paper establishes an RF-Adaboost model based on random forest feature dimension reduction and Adaboost prediction to predict

\*通讯作者。

whether the enterprise will be treated specially in year T using the financial indicators data of year T-2. The empirical results show that the classification accuracy and recall rate of RF-Adaboost model on the test set are higher than 80%. Finally, in order to verify the effect of RF-Adaboost model, this paper also uses Adaboost, LSTM neural network, RBF-SVM, Linear-SVM and kernel density naive Bayes models to conduct experiments. The results show that RF-Adaboost performs best among all models, which demonstrates the effectiveness of feature dimension reduction and the superiority of integrated algorithm.

## Keywords

Financial Distress Warning, Randomforest, Adaboost

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着经济全球化的不断发展, 我国的市场竞争环境愈发激烈。企业的财务状况是影响其未来发展的重要因素之一[1], 财务危机的发生往往是一个从异常初现到问题逐渐恶化的过程, 因此企业财务困境具有一定的可预测性。如何有效地进行企业财务困境预警从而提高企业经营质量是十分重要的, 也是国内外学者广泛关注的研究方向。

目前, 财务困境预警一般是通过建立财务指标体系来对企业的财务状况进行预测[2]。早期的预测方法技术主要为统计方法, 例如 Altman (1968)采用线性判别法建立了 Z 评分模型, 通过 Z 值大小来对企业是否陷入财务困境进行判定[3], 为了克服 Z 评分模型只适合短期预测的缺点, 改进的 ZETA 模型被提出, 预测的应用范围得到扩大[4]。Martin (1977)提出了基于逻辑回归的银行财务困境预警模型, 取得了较好的预测效果[5]。近年来, 机器学习在财务困境预警方面的应用较为热门。机器学习模型在处理大量样本和高维变量方面有着巨大的优势, 目前常用的机器学习模型有决策树、支持向量机、BP 神经网络等。崔玉娟(2011)建立了基于决策树的财务预警模型[6]。张根明和向晓骥(2007)采用了适合小样本的支持向量机算法进行上市公司财务预警研究, 实证结果表明支持向量机预测效果优于神经网络模型[7]。陈收等(2015)在支持向量机模型基础上引入了信息披露质量, 模型在企业不同发展时期的财务困境预警准确率均得到提升[8]。李晓燕(2012)建立了基于 BP 神经网络的中国上市公司财务困境预警模型, 并对神经网络的参数选取进行了研究[9]。王玉冬等(2018)将粒子群算法和果蝇优化算法与 BP 神经网络相结合, 建立了 PSO-BP 和 FOA-BP 模型进行财务危机预警[10]。

在机器学习模型应用过程中, 采用集成算法的模型的泛化效果往往优于单一模型, 常见的集成算法有 boosting, bagging, stacking 等。目前, 集成学习模型在企业财务困境预警方面的应用相对较少, 因此本文采用了 boosting 算法的典型模型 Adaboost 对上市企业进行财务困境预警研究。另外, 为了能从多个方面反映企业的财务情况, 财务困境预警模型中选取的指标一般较多, 本文创新性地利用随机森林特征重要性剔除掉贡献度低的指标来进行降维, 降低训练复杂度的同时直观反映了各财务指标在财务困境预警中的重要性, 从而挖掘出需要企业经营者重点关注的财务指标。

## 2. 模型原理

Adaboost 是 Boosting 算法的代表, 其核心思想是通过迭代训练不同的弱分类器, 构成一个强分类器。

Adaboost 进行多轮分类训练, 每轮训练会提高前一轮被弱分类器误分的样本的权重, 降低那些被正确分类的样本的权重。对于不同的弱分类器, Adaboost 采取加权投票的方法, 加大分类正确率高的弱分类器的权重, 减小分类正确率低的弱分类器的权重, 从而使得表现好的弱分类器在表决中起到较大作用, 表现差的弱分类器在表决中起到较小作用。其主要步骤如下:

1) 设有  $M$  个样本, 初始时赋予每个样本相等的权重

$$w_t(i) = \frac{1}{M}, \quad i = 1, 2, \dots, M \quad (1)$$

其中,  $w_t(i)$  表示在第  $t$  次迭代中样本  $i$  的权重。

2) 使用具有以上权重分布的  $w_t$  进行训练, 得到弱分类器  $C_t(x)$ , 其分类错误率  $e_t$  为

$$e_t = \frac{1}{M} \sum_{i=1}^M e_i \quad (2)$$

3) 根据上一轮迭代中各弱分类器的表现, 计算  $C_t(x)$  在强分类器中所占的比重  $\alpha_t$

$$\alpha_t = \frac{1}{2} \log \left( \frac{1 - e_t}{e_t} \right) \quad (3)$$

4) 更新权重分布

$$w_{t+1} = \frac{w_t}{z_t} \exp(-\alpha_t y_i C_t(x)) \quad (4)$$

$$z_t = \sum_{i=1}^M w_t \exp(-\alpha_t y_i C_t(x)) \quad (5)$$

其中,  $z_t$  为归一化因子。

5) 重复步骤(2)到(4), 直至计算出所有弱分类器的比重, 并根据比重组合得到强分类器  $S(x)$  [11]

$$S(x) = \text{sign} \left( \sum_{i=1}^M \alpha_i C_i(x) \right) \quad (6)$$

### 3. 基于 RF-Adaboost 的财务困境预警模型构建

#### 3.1. 模型构建框架

本文建立的基于 RF-Adaboost 的财务困境预警模型流程如图 1 所示。首先从营运能力、获利能力、现金流量、发展能力与偿债能力五个维度建立初始财务指标体系, 其次对上市企业样本的财务指标数据进行归一化处理和随机森林特征降维得到最终财务指标体系, 再利用 Adaboost 模型对上市企业是否会陷入财务困境进行预测。

#### 3.2. 财务指标选取

关于财务困境预警指标的选取, 根据前人的研究成果发现非财务指标的选取较为主观, 因此本文仍然选择建立财务指标体系来进行研究。杨钰晨等从现金流、偿债能力、盈利能力等方面选择多个财务指标建立了 BP 神经网络模型进行上市公司财务预警, 取得了良好的预测效果[12]。本文借鉴其对财务指标选择的分析, 从获利能力、现金流量、营运能力、发展能力、偿债能力这五个维度选取了 13 个初始财务指标, 如表 1 所示。获利能力是指企业的资金增值能力, 现金流量反映了企业获取现金的能力、财务弹性和收益质量。营运能力是指企业在受到外部市场环境约束的条件下, 通过内部人力资源和生产资料的调控而对财务目标所产生作用的大小, 可以衡量企业的资产管理效率。发展能力是指企业在维持生存的

基础上, 实现规模扩大的潜力。偿债能力是指企业到期偿还债务的能力。这五个维度可以较为全面地总结和评价企业的财务状况[13]。

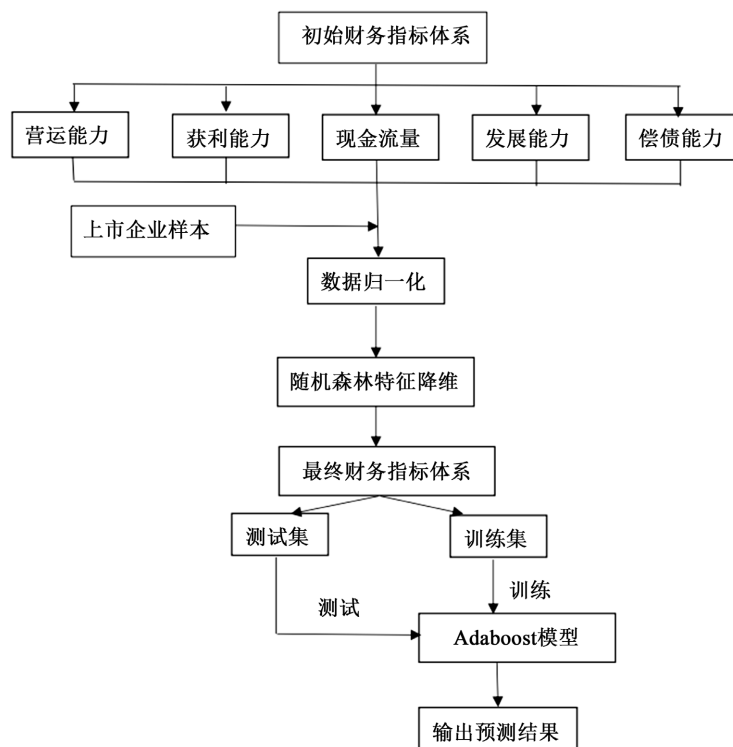


Figure 1. Modeling process  
图 1. 建模流程

Table 1. Initial financial indicators  
表 1. 初始财务指标

维度	财务指标	含义
获利能力	X1: 总资产报酬率(%)	总资产报酬率 = 息税前利润总额/资产平均总额
	X2: 净资产收益率(%)	净资产收益率 = 净利润/平均所有者权益
现金流量	X3: 营业收入现金比率(%)	营业收入现金比率 = 营业现金流量净额/营业收入
	X4: 现金回收率(%)	现金回收率 = 经营现金流量净额/平均资产总额
营运能力	X5: 流动资产周转率(次)	流动资产周转率 = 营业收入/流动资产平均占用额
	X6: 总资产周转率(次)	总资产周转率 = 营业收入/资产平均占用额
	X7: 应收账款周转率(次)	应收账款周转率 = 营业收入/应收账款平均余额
发展能力	X8: 总资产增长率(%)	总资产增长率 = 本年总资产增长额/年初资产总额
	X9: 净利润增长率(%)	净利润增长率 = 本年净利润增长额/上年净利润总额
	X10: 营业收入增长率(%)	营业收入增长率 = 本年营业收入增长额/上年营业收入总额
偿债能力	X11: 流动比率	流动比率 = 流动资产/流动负债
	X12: 资产负债率(%)	资产负债率 = 负债总额/资产总额
	X13: 权益对负债比率	权益对负债比率 = 股东权益/负债总额

### 3.3. 样本选取

根据我国证监会规定, 连续两年出现亏损的企业会被特殊处理, 股票前将被加上“ST”标记, 若其连续三年出现亏损则面临退市风险, 警告退市风险采用“\*ST”标记。因此, 企业股票被标记为 ST 或\*ST 可以作为企业陷入财务困境的标志。本文选取了 2019~2021 年沪深两市和北交所共 195 家被标记为 ST 或\*ST 的企业作为陷入财务困境的企业样本, 为了避免样本类别不均衡问题, 本文按照 1:1 的比例选取了 195 家与陷入财务困境的企业的资产规模、所在行业相近的非 ST 公司作为财务健康企业样本。

由于我国上市公司在 T 年是否会被特殊处理取决于其公布的 T-1 年的财务报告, 使用 T-1 年的财务指标数据预测企业在 T 年是否会被特殊处理没有实际意义, 因此本文采取上市公司在 T-2 年的财务指标数据构建模型来预测其在 T 年是否会被特殊处理[14]。

### 3.4. 样本数据预处理

#### 3.4.1. 数据归一化

数据归一化是建模的一项基础又十分重要的工作, 不同特征变量往往具有不同的量纲, 这会影响数据分析的结果, 为了消除量纲的影响, 需要进行归一化处理, 本文采取 min-max 标准化, 将财务指标的值映射到 [0,1] 区间内, 变换函数如下

$$x^* = \frac{x - \min}{\max - \min} \tag{7}$$

其中,  $x$  为某一财务指标的取值,  $\max$  为该财务指标的最大值,  $\min$  为该财务指标的最小值。

#### 3.4.2. 随机森林特征降维

本文采用随机森林(Random Forest, 简称 RF) OOB 误分率计算每个财务指标对企业是否会被特殊处理的分类结果的重要性。将 390 个样本用于随机森林模型训练, 得到 13 个财务指标的特征重要性如图 2 所示。由图 2 得, 贡献度排名前四的财务指标依次为: 净利润增长率、净资产收益率、总资产报酬率、总资产周转率。因此企业在经营过程中应当重点关注这些财务指标的变化。另外, 隶属于企业获利能力维度的 2 个指标总资产报酬率、净资产收益率, 隶属于企业发展能力维度的 3 个指标总资产增长率、净利润增长率、

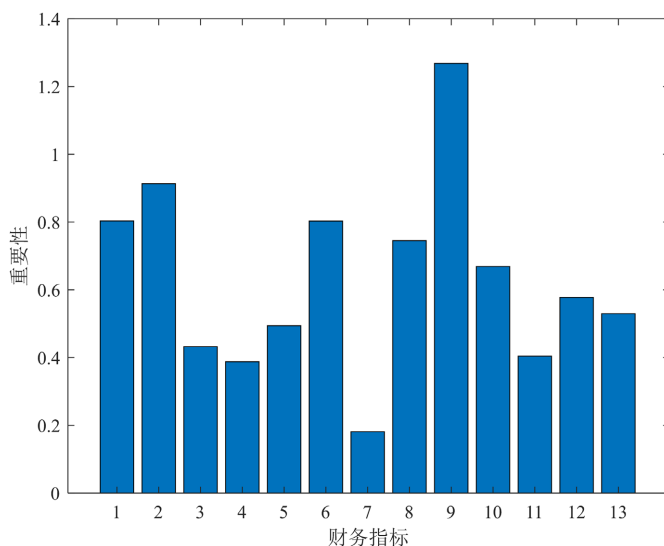


Figure 2. Importance of financial indicators  
图 2. 财务指标重要性

营业收入增长率的特征重要性都较高。连续 2~3 年出现亏损的企业将会面临被特殊处理的风险, 企业能否获利直接反映了其是否有足够的资金维持自身发展。企业的发展能力可以反映其在未来的财务状况走势, 发展能力强的企业即使暂时存在轻微财务困难, 但其具有转亏为盈的潜力而免于陷入财务困境。所以, 企业的获利能力和发展能力在判断企业是否会陷入财务困境中有较大的参考价值。

本文将重要性低于 0.6 的财务指标予以剔除, 最终保留的 6 个财务指标为: 总资产报酬率、净资产收益率、总资产周转率、总资产增长率、净利润增长率、营业收入增长率。

## 4. 模型实验结果分析

为了便于模型进行训练和预测, 本文将未被特殊处理, 即财务健康的企业归为类别 0, 将被特殊处理, 即陷入财务困境的企业归为类别 1。在衡量机器学习模型的性能时, 应重点关注模型在测试集上的表现来对模型的泛化能力进行评估。

### 4.1. 模型评价指标

#### 4.1.1. 分类正确率

模型的分类正确率是重要的衡量模型性能的指标, 能够反映出模型对企业是否会陷入财务困境的判别能力。

$$\text{Accuracy} = \frac{x_{11} + x_{00}}{x_{11} + x_{10} + x_{01} + x_{00}} \quad (8)$$

Accuracy 为分类正确率,  $x_{ij}$  为实际类别为  $i$ , 预测类别为  $j$  ( $i, j = 0, 1$ ) 的样本个数, 测试集分类正确率越高, 表明模型的泛化能力越强。

#### 4.1.2. 召回率

为了起到预警作用, 将实际情况为陷入财务困境的企业正确预测显得尤为重要。如果预测错误, 可能导致已经出现财务危机信号的企业对未来抱有错误预期, 未能意识到问题的严重性, 从而无法及时采取措施而陷入财务困境。因此除了分类正确率, 本文引入召回率作为第二个模型评价指标。

$$\text{Recall} = \frac{x_{11}}{x_{11} + x_{10}} \quad (9)$$

Recall 为召回率,  $x_{ij}$  为实际类别为  $i$ , 预测类别为  $j$  ( $i, j = 0, 1$ ) 的样本个数。

### 4.2. 模型预测结果

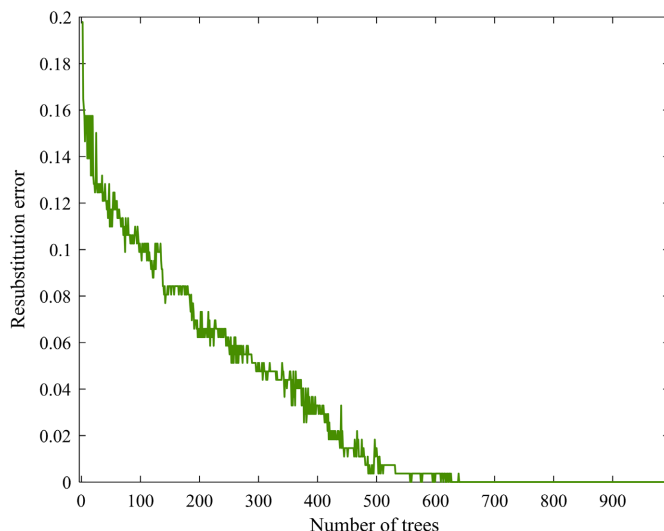
本文按照训练集样本数:测试集样本数 = 7:3 的比例, 将 273 个样本作为训练集, 将 117 个样本作为测试集, 进行模型的训练和预测。Adaboost 的再代入误差(Resubstitution error)是其在训练集上的分类错误率, 随着子分类树的数目的增加, 再代入误差数值不断减小, 并趋于相对平稳的状态, 如图 3 所示。本文选择 800 作为 RF-Adaboost 模型的树的数目, 其训练和预测结果如表 2 所示。RF-Adaboost 模型在训练集和测试集上的表现都较好, 测试集的分类正确率达到 84.62%, 测试集召回率达到 80.4%。这说明模型不仅对企业是否会陷入财务困境取得了理想的预测效果, 还具有对实际情况为陷入财务困境的企业进行正确识别的能力, 具有较好的预警作用。

### 4.3. 模型预测结果验证

#### 4.3.1. RF-Adaboost 与 Adaboost 模型对比

为了验证随机森林特征重要性进行特征降维是否有效, 本文比较了直接使用初始 13 个财务指标的

Adaboost 模型和使用特征降维的 RF-Adaboost 模型的预测效果, 如表 3 所示。由表 3 得, RF-Adaboost 在测试集上的分类正确率和召回率均高于 Adaboost 模型, 说明特征降维是有效的。



**Figure 3.** Change of resubstitution error  
**图 3.** 再代入误差变化

**Table 2.** Results of RF-Adaboost model  
**表 2.** RF-Adaboost 模型结果

评价指标	数值
训练集分类正确率	98.53%
训练集召回率	97.80%
测试集分类正确率	84.62%
测试集召回率	80.40%

**Table 3.** Comparison of RF-Adaboost and Adaboost  
**表 3.** RF-Adaboost 与 Adaboost 对比

模型	评价指标	评价指标数值
Adaboost	测试集分类正确率	78.64%
	测试集召回率	76.00%
RF-Adaboost	测试集分类正确率	84.62%
	测试集召回率	80.40%

### 4.3.2. 其他机器学习模型对比

为了进一步验证模型的有效性和优越性, 本文在随机森林特征降维的基础上使用了多个其它机器学习模型进行预测效果对比。在企业财务困境预警领域, 朴素贝叶斯和支持向量机是比较常见的机器学习模型。朴素贝叶斯方法中对变量的独立性和分布有较高的要求, 因此本文采用基于核密度估计这一非参数概率估计方法的朴素贝叶斯模型进行训练和预测。核函数是支持向量机模型的最重要的参数之一, 核

函数的选择对模型的性能有很大影响,因此本文选择了 RBF 和 Linear 两种核函数分别建立支持向量机模型进行训练和仿真。近年来,一些学者开始将神经网络用于财务困境预警,相对于传统神经网络,以 RNN、LSTM 等为代表的神经网络可以更好地捕捉变量间的复杂非线性关系,LSTM 神经网络在 RNN 神经网络的基础上引入了输入门、遗忘门和输出门,通过门控状态来控制过往数据向未来的传输状态,实现重要信息的长期记忆和不重要数据的及时遗忘。因此,本文选择 LSTM 神经网络进行训练与预测。各模型的测试集表现如表 4 所示。

由表 4 得出,RF-Adaboost 的测试集分类正确率和召回率均高于其他模型,其泛化能力最好,这说明了相较于单一模型,集成学习模型在分类效果上的优越性。

**Table 4.** Comparison of different machine learning models

**表 4.** 不同机器学习模型效果对比

模型	测试集分类正确率	测试集召回率
RF-Adaboost	84.62%	80.40%
LSTM 神经网络	81.19%	78.00%
RBF-SVM	74.36%	66.68%
Linear-SVM	75.21%	71.60%
基于核密度估计的朴素贝叶斯	78.64%	37.55%

## 5. 研究结论与展望

本文通过构建基于特征降维和集成学习理论的 RF-Adaboost 模型,对我国上市企业进行了财务困境预警研究。本文创新性地根据随机森林特征重要性进行降维,直观反映了各财务指标在财务困境预警中的重要性,从而挖掘出需要企业经营者重点关注的财务指标。研究表明,净利润增长率、净资产收益率、总资产报酬率、总资产周转率这 4 个指标对企业是否会陷入财务困境具有较强的指向作用。同时,企业应当注重自身的获利能力和发展能力,不仅要立足当下,还要力争成为“潜力股”。

其次,为了验证 RF-Adaboost 模型的效果,本文还使用了 Adaboost 模型,LSTM 神经网络、RBF-SVM、Linear-SVM、基于核密度估计的朴素贝叶斯模型进行了实验,研究表明 RF-Adaboost 在所有模型中表现最好,测试集的分类正确率和召回率都达到 80% 以上,说明了特征降维的有效性和集成算法的优越性。

本文的研究成果为一般企业提供了可操作性强的财务困境预警方法,但在实际情况中,除了财务情况这一主要因素,导致企业陷入财务困境的因素还有很多。例如,企业的管理机制不当可能会导致企业经营不善,企业领导者在重大决策上的失误可能造成企业的巨大损失从而陷入困境等。不同企业的情况不同,要想精准地对企业的财务状况进行预测,还需有针对性地对企业进行全面的调查和分析,因此在实际应用中,如何制定个性化的企业财务困境预警方案是未来需要研究的问题。

## 参考文献

- [1] 葛浩. 基于 CART 决策树模型的上市企业财务风险预警研究[D]. 上海: 上海财经大学, 2020.
- [2] 王昱, 杨珊珊. 考虑多维效率的上市公司财务困境预警研究[J]. 中国管理科学, 2021, 29(2): 32-41.
- [3] Altman, E.I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23, 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>



- [4] Altman, E.I., Haldeman, R.G. and Narayanan, P. (1977) ZETA Analysis a New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking & Finance*, **1**, 29-54. [https://doi.org/10.1016/0378-4266\(77\)90017-6](https://doi.org/10.1016/0378-4266(77)90017-6)
- [5] Martin, D. (1977) Early Warning of Bank Failure: A Logit Regression Approach. *Journal of Banking & Finance*, **1**, 249-276. [https://doi.org/10.1016/0378-4266\(77\)90022-X](https://doi.org/10.1016/0378-4266(77)90022-X)
- [6] 崔玉娟. 基于决策树的财务危机预警研究[J]. 财会通讯, 2011, 529(17): 147-148.
- [7] 张根明, 向晓骥. 基于支持向量机的上市公司财务预警模型研究[J]. 科技管理研究, 2007, 170(4): 234-235+242.
- [8] 陈收, 赖柏纯, 杨宽. 基于信息披露质量的财务预警研究[J]. 经济与管理研究, 2015, 36(8): 122-129.
- [9] 李晓燕. 中国上市公司财务危机预警研究——基于 BP 神经网络的分析[J]. 西南民族大学学报(人文社会科学版), 2012, 33(12): 156-160.
- [10] 王玉冬, 王迪, 王珊珊. 基于 PSO-BP 和 FOA-BP 神经网络的财务危机预警模型比较[J]. 统计与决策, 2018, 34(15): 177-179.
- [11] 宁昱博, 张玉军. LSTM-Adaboost 股价预测模型[J]. 辽宁科技大学学报, 2019, 42(5): 383-388.
- [12] 杨钰晨, 丁元耀. 基于 BP 神经网络的上市公司财务预警模型研究[J]. 生产力研究, 2021, 346(5): 155-160.
- [13] 陈玉菁, 宋良荣. 财务管理[M]. 第 4 版. 北京: 清华大学出版社, 2016: 279-294.
- [14] 谭媛元, 陈建英, 孙健. 基于 CNN 的上市公司财务危机预警研究[J]. 西南师范大学学报(自然科学版), 2021, 46(5): 73-80.