

基于机器学习方法的中国上市公司舞弊研究

于李胜^{a,b}, 郑天宇^b, 滕传浩^b

(厦门大学 a. 会计发展研究中心; b. 管理学院 福建 厦门 361005)

摘要:数字时代的来临使利用机器学习识别财务舞弊成为研究的热点。在采用原始财务数据的基础上,引入财务比率、公司治理指标、审计指标和我国资本市场特殊指标,并以 Logistic 模型为评价基准,分别运用决策树、随机森林、Adaboost 决策树和支持向量机(SVM)模型进行机器学习分析,并利用采样的方式降低样本不平衡性,以召回率(recall)为评价各模型的标准,综合运用准确率(accuracy)、召回率和 AUC 判断模型和数据的优劣。研究发现:加入财务比率、审计指标和我国资本市场特殊要素指标的模型能够得到较优的识别效果,而公司治理指标并不能提高模型的舞弊识别能力;与其他模型相比,随机森林模型和 Adaboost-决策树模型具有更好的舞弊识别效果。

关键词:数字经济;机器学习;舞弊识别;上市公司

中图分类号: F239.1 文献标识码: A 文章编号: 0438-0460(2023)02-0044-12

一、引言

为防范化解金融风险,维护金融市场秩序和完善资本市场建设,我国监管部门必须对上市公司财务舞弊等违法犯罪行为采取“零容忍”态度。上市公司舞弊行为不仅损害了投资者利益,而且降低了资本市场有效性。数字时代的来临使财务舞弊呈现出了新的特点和趋势,随着计算机技术和相关硬件设施的不断发展,使用机器学习方法识别财务舞弊成为研究的热点话题。在国外的相关研究中,已有学者利用机器学习得到较好的舞弊识别和预测效果^①,国内相关研究也借鉴国外的研究方法^②。与美国等发达国家资本市场相比,我国的资本市场具有散户比重大和单日交易量高等特征,这些特征表明仅仅利用国外的方法和数据研究我国的资本市场舞弊问题会存在较大偏差。因此,如何结合我国国情设计符合中国制度背景的机器学习财务舞弊识别模型是一个值得探索的研究问题。

收稿日期: 2022-10-06

基金项目: 国家自然科学基金面上项目“信息披露对实体经济发展的影响机制研究”(71972161);国家自然科学基金面上项目“会计准则视角下银行系统性风险影响机制研究”(71972162);校长基金创新团队项目(20220191087)

作者简介: 于李胜,男,山东海阳人,厦门大学会计发展研究中心/管理学院教授、博士生导师,管理学博士;郑天宇,男,浙江奉化人,厦门大学管理学院博士研究生;滕传浩,男,山东济南人,厦门大学管理学院博士研究生。

① Bao, Y., Ke, B., Li, B., Yu, Y. J., Zhang, J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research*, 2020, 58(1), pp. 199–235; Cecchini, M., Aytug, H., Koehler, G. J., Pathak, P., “Detecting management fraud in public companies”, *Management Science*, 2010, 56(7), pp. 1146–1160.

② 王泽霞、郜鼎、李冬艳《基于 BP-LVQ 的组合神经网络舞弊风险识别模型研究——来自中国舞弊上市公司的经验证据》,《生产力研究》2017 年第 12 期; 黄志刚、刘佳进、林朝颖《基于机器学习的上市公司财报舞弊识别前沿方法比较研究》,《系统科学与数学》2020 年第 10 期。

本文在采用原始财务数据的基础上,逐步引入财务比率、公司治理指标、审计指标和我国资本市场特殊指标,并以 Logistic 模型为评价基准(M-Score、F-Score 和 C-Score),分别运用决策树、随机森林、Adaboost-决策树和支持向量机(SVM) 模型进行机器学习分析,利用过采样的方式降低样本不平衡性,以召回率(recall) 为评价各模型的标准,同时兼顾成本收益原则,综合运用准确率(accuracy)、召回率和 AUC 判断模型和数据的优劣。研究发现: 在采用原始财务数据基础之上,加入财务比率、审计指标和我国资本市场特殊要素指标的模型能够得到较优的识别效果,而公司治理指标并不能提高模型的舞弊识别能力;与其他模型相比,随机森林模型和 Adaboost-决策树模型具有更好的舞弊识别效果,准确率分别为 62% 和 64%,且召回率为 64% 和 62%。同时,样本外的识别结果表明本文所采用的模型具有较高的稳健性。

本文的创新性与潜在贡献主要体现在以下方面:一方面,本文丰富了公司舞弊识别领域的文献,现有的公司舞弊识别研究主要是基于传统财务数据^①,忽视了我国资本市场特殊指标对舞弊识别的预测作用。本文在财务数据的基础上,逐渐引入公司治理指标、审计指标和我国资本市场特殊指标,使得舞弊识别模型更适合我国股票市场背景和企业特征,模型具有更高的适用价值。另一方面,本文所提出的舞弊识别模型有助于提高资本市场的有效性,维护投资者利益。在全面实行股票发行注册制的背景下,监管机构和投资者更多依靠信息披露对企业进行监督,^②而本文利用企业披露的多维信息进行舞弊识别,为监管部门和投资者提供了更有效的舞弊识别工具。

二、文献回顾

公司舞弊行为已经引起了国内外学者的广泛关注,随着企业舞弊行为的研究不断深入,大量文献集中讨论了企业舞弊的识别和预测。Persons 利用公开财务信息进行舞弊识别,依据 Logistic 回归筛选出四个与舞弊相关的指标,企业规模小、流动资产比重高、杠杆率高而资本周转率低的企业更可能舞弊。^③ Beneish 利用 Probit 模型提出了对舞弊预测的概率分析,建立了包括应收账款周转率指数、毛利率指数、资产质量指数、销售增长指数和应计项目比例共五个财务指标的舞弊预测模型。^④ 另外,也有部分学者利用财务数据预测公司舞弊行为。Beneish 利用 1982—1992 年被美国证监会(SEC) 查处的 74 家财务造假公司为观察样本,构建了以 8 项财务指标为基础的 M-Score 舞弊预测模型。^⑤ Dechow 等在 M-Score 基础上建立了 F-Score 舞弊预测模型,从应计项目、财务指标、非财务指标、表外业务和市场信息五个方面全面检验了造假公司的特征。^⑥ 随着计算机技术的不断发展,个人计算机的硬件设施和算力不断增强,利用计算机技术进行舞弊识别和预测成为学者研究的热点话题。Cecchini 等利用支持向量机(SVM) 方法建立了财务舞弊识别模型,识别和预测结果表明其准确率达到 80%。^⑦ Bao 等以财务报表原始数据作为数据来源,采用集成学习的方法构

^① Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

^② 陈淮、顾连书《我国股票发行注册制的制度条件及其政策研究》,《上海财经大学学报(哲学社会科学版) 》2012 年第 2 期。

^③ Persons , O. S., “Using financial statement data to identify factors associated with fraudulent financial reporting”, *Journal of Applied Business Research* , 1995 , 11(3) , pp. 38–46.

^④ Beneish , M. D., “Detecting GAAP violation: Implications for assessing earnings management among firms with extreme financial performance”, *Journal of accounting and public policy* , 1997 , 16(3) , pp. 271–309.

^⑤ Beneish , M. D., “The detection of earnings manipulation”, *Financial Analysts Journal* , 1999 , 55(5) , pp. 24–36.

^⑥ Dechow , P. M. , Ge , W. , Larson , C. R. , Sloan , R. G. , “Predicting material accounting misstatements”, *Contemporary accounting research* , 2011 , 28(1) , pp. 17–82.

^⑦ Cecchini , M. , Aytug , H. , Koehler , G. J. , Pathak , P., “Detecting management fraud in public companies”, *Management Science* , 2010 , 56(7) , pp. 1146–1160.

建舞弊识别模型,根据结果显示该模型优于支持向量机和逻辑回归模型。^①

钱萍和罗玫参考 M-Score 和 F-Score 的构建,并考虑我国资本市场的特殊性提出了 C-Score 模型,经营应计项指数、现金销售率、股票换手波动率、股权集中度、机构投资者持股比率、其他应收款比例、公司当年是否处于亏损状况、是否进行了再融资和股市是否处于“熊市”等变量显著地反映了我国财务造假公司的特征。^② 洪文洲等利用 Logistic 模型对我国上市公司进行了舞弊识别和预测,研究结果表明折旧率提升、应收账款周转期延长、股权分散等现象是发生财务舞弊的重要特征。^③ 在机器学习算法方面,于彪和陈思凤在 SVM 方法的基础上,采用了样本不平衡程度较低的数据,其模型准确率达到 98%。^④ 高宇采用朴素贝叶斯网络,利用 24 个财务指标建立舞弊识别模型,其识别准确率为 70%。^⑤ 黄志刚等比较了多种机器学习方法在中国资本市场的优劣,结果表明随机森林、支持向量机及神经网络等方法优于其他方法,随机森林在测试集中表现最佳。^⑥

综上所述,关于企业舞弊的识别和预测研究已经较为成熟,提出了切实有效的识别模型。随着计算机技术的发展,舞弊识别的研究从原来的 Logistic 回归和构建舞弊特征指数进入到利用机器学习方法识别和预测的阶段。相关学者利用财务报告原始数据,并根据以往舞弊影响因素的文献,加入了公司治理和审计的相关指标,优化了模型的数据来源和准确性。但是,现有的舞弊识别文献忽略了我国资本市场的特殊性,过多关注于优化模型方法,我国的上市公司存在一些与其他国家不同的特征,比如商誉爆雷、关联交易金额高以及股权质押比例高,这些特点对企业的资产质量、业绩持续性和经营风险等方面带来不同的影响,进而会影响到公司舞弊的压力和动机。且以往的研究中主要运用准确率和精确率等评价指标,主要考虑识别的成本原则却忽视了召回率。我们认为召回率代表正确识别所有发生舞弊公司的比例,反映了效率原则,也应纳入模型的评价标准中。

三、模型构建与数据选择

(一) 模型构建

机器学习的方法较多,本文将运用决策树、随机森林、Adaboost-决策树和支持向量机(SVM)这四种方法进行舞弊识别,比较不同模型在中国资本市场中的优劣以及不同输入值的识别效果。

为了评价模型的优劣,我们需要引入相关的评价指标进行分析。通常来说,模型的评价指标有准确率(accuracy)、精确率(precision)、召回率(recall)、接收者操作特征曲线下面积(AUC)等评价指标。于彪和陈思凤以准确率为评价标准进行分析^⑦,王泽霞等以精确率为标准进行神经网络分析^⑧,Cecchini 等运用 ROC 曲线评价了该篇论文所提出模型的优劣性^⑨。Bao 等运用 AUC 和归一

^① Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

^② 钱萍、罗玫《中国上市公司财务造假预测模型》,《会计研究》2015 年第 7 期。

^③ 洪文洲、王旭霞、冯海旗《基于 Logistic 回归模型的上市公司财务报告舞弊识别研究》,《中国管理科学》2014 年第 S1 期。

^④ 于彪、陈思凤《会计信息失真识别的成本控制型支持向量机模型》,《商场现代化》2011 年第 6 期。

^⑤ 高宇《基于贝叶斯网络的企业财务报告舞弊识别研究》,《经营与管理》2018 年第 11 期。

^⑥ 黄志刚、刘佳进、林朝颖《基于机器学习的上市公司财报舞弊识别前沿方法比较研究》,《系统科学与数学》2020 年第 10 期。

^⑦ 于彪、陈思凤《会计信息失真识别的成本控制型支持向量机模型》,《商场现代化》2011 年第 6 期。

^⑧ 王泽霞、郜鼎、李冬艳《基于 BP-LVQ 的组合神经网络舞弊风险识别模型研究——来自中国舞弊上市公司的经验证据》,《生产力研究》2017 年第 12 期。

^⑨ Cecchini , M. , Aytug , H. , Koehler , G. J. , Pathak , P., “Detecting management fraud in public companies”, *Management Science* , 2010 , 56(7) , pp. 1146–1160.

化折损累计增益评价模型的识别和预测能力。^① 本文将重点关注模型的召回率,同时在考虑成本收益的前提下,选择适合我国资本市场的最优模型。

具体而言,对于二分类的问题,样本结果可根据真实类别和预测结果划分为一个 2×2 的混淆矩阵,四种结果分别为真阳(True Positive)、假阳(False Positive)、真阴(True Negative) 和假阴(False Negative),分别简称为 TP、FP、TN 和 FN。

其中准确率表示分类正确的样本占总样本的比例,计算公式如下:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

精确率表示真阳性样本占预测为真样本的比例,计算公式如下:

$$\text{Precision}(1) = \frac{TP}{TP + FP}$$

类似地,真阴性样本占预测为假样本的比例如下所示:

$$\text{Precision}(0) = \frac{TN}{TN + FN}$$

召回率表示真阳性样本占实际为真样本的比例,计算公式如下:

$$\text{Recall}(1) = \frac{TP}{TP + FN}$$

类似地,真阴性样本占实际为假样本的比例如下所示:

$$\text{Recall}(0) = \frac{TN}{TN + FP}$$

从上面的几个指标来看,准确率是一个对模型分类结果全面考量的指标。但是,由于样本不平衡的存在,将样本数据全部识别为非舞弊数据也具有较高的准确率,因此仅使用准确率作为评价指标不能合理体现出模型的优劣。精确率也具有相同的问题,若将样本数据全部识别为某一类别,该类别的精确率等于1。然而召回率表示对真实类别分类的正确性,更适合作为评价模型优劣的指标。Recall(1)的经济意义更符合本文的目的,Recall(1)表示模型能够正确识别出所有发生舞弊公司的比例,即模型识别和预测舞弊的能力。基于监管者的角度,我们首先要保证尽可能多地找到发生舞弊的公司,减少市场中存在舞弊公司的可能性,保护投资者利益。其次才是保证模型的准确率,减少模型错误识别非舞弊为舞弊的比例,降低监管工作成本。为此,本文的主要衡量指标为召回率,同时基于成本收益的原则,手动调整模型参数,保证各模型的准确率处在同一水平,利用召回率和准确率综合评价模型和数据的优劣。

(二) 样本数据选择

本文的样本数据选择首先参考了现有的利用机器学习进行财务舞弊识别的文献,选取可以从财务报告中直接获取的原始财务数据,再加入相关财务指标,对原始财务数据进行补充,使得不同公司之间的财务数据更加可比;之后查阅借鉴相关影响财务舞弊的因素的研究,分别加入了公司治理指标和审计指标,以提升机器学习模型的识别和预测的准确性;最后,因为本文利用我国资本市场数据进行研究,与美国样本数据相比,我国上市公司的特征具有一定的差别,如商誉和股权质押等,所以需要在样本中加入我国资本市场和上市公司的特殊要素。

1. 原始财务数据

原始财务数据是指直接列示在企业财务报表上的金额,比如资产负债表的账户余额和利润表的当期发生额。原始财务数据是会计系统中最基础的组成部分,利用这些数据可以避免人为对数

^① Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

据的加工所产生的噪音,判断原始数据能否直接用于舞弊识别。Bao 等认为利用原始财务数据进行舞弊识别主要是由于关于会计舞弊驱动因素的现有理论很可能是不完整的,会计舞弊是秘密进行的,而且设计得很难被发现。^①因此,将原始会计数据转换成财务比率可能意味着有用的预测信息的损失。本文借鉴了现有文献中所选取的原始财务数据,结合数据可获得性,筛选出共计 18 个原始财务数据。

2. 财务比率

在传统的财务舞弊识别的文献中,研究人员一般会选用财务比率作为研究对象来判别上市公司舞弊行为。^②当公司有动机采取舞弊行为时,这些已有的研究可以为投资者提供合理的预测,即利用现有的财务比率可以提供一个成本较低的舞弊识别方法。另外,若只采用原始财务数据或财务比率数据进行研究,所选取的样本很可能是不完整的。利用财务比率数据作为原始财务数据的补充能够更好地提升机器学习的预测能力,增强舞弊的识别效率。本文参考以往文献,结合我国资本市场公司的特征和舞弊案例,从偿债能力、经营能力、盈利能力和现金流分析四个方面选取了 13 个财务比率。

偿债能力衡量了企业使用其自有资产偿还债务的能力,它是反映企业财务状况和经营风险的重要指标。Fanning 和 Cogger 研究认为,当企业面临巨大的债务压力时,管理层发生舞弊行为的可能性增强。^③一般而言具有高负债率同时偿债能力较弱的公司有较高的舞弊动机,因此,本文选取了流动比率、现金比率、利息保障倍数和资产负债率作为衡量企业偿债能力的指标。

经营能力衡量企业对生产经营活动的决策和控制权力,也能够体现出企业在供应链中所处的地位。Persons 研究认为舞弊公司与非舞弊公司在经营资产上存在差异,舞弊公司的流动资产的比例更高,应收账款和存货比例较高。^④同时舞弊企业采用虚构交易和虚增收入的方式调整利润金额,在这种情况下应收账款周转率、存货周转率以及总资产周转率都会低于一般企业。另外,企业的综合经营能力可以由企业的实际所得税率体现,企业的每次交易行为都需要交税,税收对企业产生额外的成本,企业虚构交易只会增加企业成本而没有利润的增长,因此舞弊公司存在着所得税费用与利润不配比的情况。本文选取了应收账款周转率、存货周转率、总资产周转率和所得税率衡量企业的经营能力。

盈利能力体现企业资本增值的能力,Erickson 等研究认为公司财务绩效和舞弊行为之间存在负相关关系,绩效差的公司往往具有高舞弊动机。^⑤因此,本文选取总资产收益率(ROA)和营业利润率衡量企业的盈利能力。

本文选取了现金流分析指标作为上述三类指标的补充,利用 ROA 和营业利润率可以有效衡量企业的盈利能力,但在权责发生制下,企业为实现收入额增加会采用赊销的方式,仅用上述两个指标无法衡量企业的盈利质量,即企业通过销售活动收回多少现金。曹德芳和刘柏池认为公司通过虚拟交易的方式增加公司利润,然而经营性现金流不会随之增加。^⑥本文选择净利润现金净含量、

^① Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

^② Dechow , P. M. , Ge , W. , Larson , C. R. , Sloan , R. G., “Predicting material accounting misstatements” , *Contemporary accounting research* , 2011 , 28(1) , pp. 17–82.

^③ Fanning , K. M. , Cogger , K. O., “Neural network detection of management fraud using published financial data” , *Intelligent Systems in Accounting , Finance & Management* , 1998 , 7(1) , pp. 21–41.

^④ Persons , O. S., “Using financial statement data to identify factors associated with fraudulent financial reporting” , *Journal of Applied Business Research* , 1995 , 11(3) , pp. 38–46.

^⑤ Erickson , M. , Hanlon , M. , Maydew , E. L., “Is there a link between executive equity incentives and accounting fraud?” , *Journal of accounting research* , 2006 , 44(1) , pp. 113–143.

^⑥ 曹德芳、刘柏池《SVM 财务欺诈识别模型》,《东北大学学报(自然科学版)》2019 年第 2 期。

现金流利息保障倍数和应计项目这三项指标进行分析,完善财务比率的分析结果。

3. 公司治理指标

企业的经营业绩只能体现出企业的经营结果而无法体现出企业在经营过程中出现的问题,非财务指标也引起了学者的广泛关注。公司治理体现了公司进行管理和控制的水平,表现出公司各方面参与者在公司内部权力分配和责任制定。管理层是公司决策的核心部门,管理层的能力与特征直接影响着公司行为。已有文章表明,董事会和管理层与公司舞弊行为密切相关,^①加入相关数据能够使得利用机器学习方法具有更好的泛化能力。另外,参考现有机器学习的相关文献,部分文献中已经加入公司治理相关指标进行舞弊识别研究,^②且能够得到较高的实证结果。因此本文选取管理层持股数量、董事人数、董事长与总经理兼任情况、董事、监事及高管年薪总额以及委员会设立总数五个指标衡量公司治理水平。

4. 审计指标

当企业公司治理不善和内部控制失效时,外部审计就成为保证信息披露质量的关键要素。利益相关者期望得到真实可靠的财务报告从而做出相关决策,审计质量的高低代表着企业舞弊风险程度。首先,审计费用的高低及其确定方式会影响审计质量。李明辉和沈真真认为,过高的异常收费与审计质量显著负相关,而过低的异常收费与审计质量之间的关系不显著。^③其次,会计师事务所变更也可能是出现舞弊行为的信号,若上市公司当期出现舞弊行为并被审计师发现,上市公司需要遵从审计师意见对财务报表进行修改,否则审计师将出示非标准审计意见,非标准审计意见将会对公司股价产生负面影响。公司若不想对财务数据进行修改且继续使用现有报告,则会选择解雇当前会计师事务所并聘任能够听取管理层意见的会计师事务所,马晨等发现在财务重述后的一年内,事务所被解聘的频率在财务重述公司中要高于未发生财务重述的公司,可见财务报告的真实性与事务所变更之间存在相关关系。^④另外,部分研究认为四大会计师事务所是审计质量的替代变量,漆江娜等经四大会计师事务所审计的公司的操控性应计利润额略低于本土事务所审计的公司,可见事务所选择会影响企业的盈余管理行为。^⑤因此,本文选取审计费用、是否变更会计师事务所、审计意见和审计师是否来自国际四大会计师事务所作为审计指标输入值。

5. 我国资本市场和上市公司特殊要素

我国资本市场经过三十年左右的发展已经取得了较大成功,资源配置水平日趋成熟,但与美国等发展起步时间早的其他世界主流资本市场相比,我国资本市场还是具有较大差异。首先,我国的资本市场中以散户投资者为主,市场参与者的交易行为与美国资本市场不同。同时,我国资本市场处于改革转型阶段,在全面实行股票发行注册制的背景下,公司上市制度从核准制转变到注册制后,企业上市难度下降,监管责任后移,主管机构只负责审查发行人是否履行信息披露义务,不对拟发行股票投资价值进行判断。在这种情况下,监管行为由事前转向事中和事后,更多依靠市场行为识别财务舞弊。可见,我国资本市场正处于一个发展与改革并行的阶段,也有很多与其他发达资本

^① Beasley , M. S., "An empirical analysis of the relation between the board of director composition and financial statement fraud", *Accounting review* , 1996 , pp. 443–465; Chen , G. , Firth , M. , Gao , D. N. , Rui , O. M., "Ownership structure , corporate governance , and fraud: Evidence from China" , *Journal of corporate finance* , 2006 , 12(3) , pp. 424–448.

^② Dechow , P. M. , Sloan , R. G. , Sweeney , A. P., "Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC" , *Contemporary accounting research* , 1996 , 13(1) , pp. 1–36; 张健、刘斌、吴先聪《财务舞弊、家族控制与上市公司高管更替》,《管理工程学报》2015年第2期。

^③ 李明辉、沈真真《异常收费与审计质量:来自中国资本市场的经验证据》,《管理工程学报》2016年第2期。

^④ 马晨、张俊瑞、杨蓓《财务重述对会计师事务所解聘的影响研究》,《会计研究》2016年第5期。

^⑤ 漆江娜、陈慧霖、张阳《事务所规模·品牌·价格与审计质量——国际“四大”中国审计市场收费与质量研究》,《会计研究》2004年第3期。

市场国家不同的问题,仅利用国外研究的方法探究我国资本市场问题可能会造成研究偏差。

但从现有文献中可以看出,利用机器学习方法进行我国上市公司舞弊的识别的研究都是利用国外研究的方法和数据^①,忽略了我国资本市场和上市公司的特殊性。本文将引入相关的特征数据,关注我国资本市场和上市公司的独特性,这样进行数据分析才更适应我国上市公司财务舞弊的情景。企业若希望采用虚增收入的方式实现舞弊行为,构建的交易合同大多是存在与关联方之间。郝玉贵和刘李晓认为,关联方交易舞弊风险因素具有多样性,上市公司关联方交易舞弊的动机在于提升股价,实现解禁套现最大化的目的。^②冯惠琴认为,关联企业之间为了达到粉饰报表、调节利润的目的,可能存在关联购销,资产重组,费用分担,资产租赁与委托经营和资金占用与资金往来等问题。^③大股东资金占用问题是我国资本市场的一个重要弊病,其他应收款由于其核算复杂,便于对一些性质模糊的业务进行账务处理,实质为大股东资金占用的相关账目便会记录在该科目中,从而导致其他应收款是近几年监管层重点管理的问题。郑国坚等利用其他应收款数据实证发现面临财务困境时,大股东对上市公司的非法资金占用行为异常明显,显示其强烈的掏空动机,发生舞弊的规律更高。^④姜国华和岳衡认为,大股东资金占用和上市公司未来年度的盈利能力呈显著负相关关系,表明市场对大股东资金占用行为具有负向认知。^⑤

大股东掏空的另一个表现为股权质押。郑国坚等认为,用股权质押衡量大股东的财务约束及其掏空倾向,对大股东的掏空行为具有较强的解释力。^⑥当大股东面临严重财务约束(股权被质押、冻结)时,更容易对上市公司进行占款。我国上市公司中普遍存在股权质押行为,存在控股股东股权质押的公司真实盈余管理水平更高,增加舞弊风险。^⑦吕晓亮认为控股股东股权质押提高了公司违规后被稽查的概率,降低了公司的违规倾向,有利于监管层进行舞弊识别。^⑧

当企业面临巨大的业绩压力时,管理层可以采取并购的方式,将一家资产质量较好盈利水平较高的公司合并入公司的财务报表中,这使得合并财务报表比母公司财务报表具有更好的经营成果,实现业绩反转。然而,在财务压力较大的情况下贸然发起收购行为会进一步放大企业的财务风险(如杠杆收购),存在资金链断裂的可能性。同时收购方需要与被收购方签订业绩承诺,李晶晶等认为,并购业绩承诺使得上市公司未来股价暴跌风险加重,业绩补偿承诺引致代理冲突变异并激发相关利益方的机会主义行为。^⑨机会主义行为便会增加企业舞弊风险,可见并购与业绩承诺已经成为审计师重点关注的问题,其中隐藏着业绩“变脸”的重大风险。

另外,随着并购交易在我国资本市场中的数量不断增加,商誉在上市公司财务报表中所占比重越来越大。李晶晶等研究发现商誉资产占上市公司净资产比例最高达5.22%,同时并购业绩承诺使得并购溢价和商誉确认金额更高,加大了未来商誉减值概率,进一步分析发现并购业绩承诺制度降低了商誉的价值相关性。^⑩可见商誉是一个在我国上市公司广泛存在而且对市场产生负面影响

^① 黄志刚、刘佳进、林朝颖《基于机器学习的上市公司财报舞弊识别前沿方法比较研究》,《系统科学与数学》2020年第10期;曹德芳、刘柏池《SVM财务欺诈识别模型》,《东北大学学报(自然科学版)》2019年第2期。

^② 郝玉贵、刘李晓《关联方交易舞弊风险内部控制与审计——基于紫鑫药业案例的研究》,《审计与经济研究》2012年第4期。

^③ 冯惠琴《基于关联方交易的会计舞弊研究》,《山西财经大学学报》2011年第S2期。

^④ 郑国坚、林东杰、张飞达《大股东财务困境、掏空与公司治理的有效性——来自大股东财务数据的证据》,《管理世界》2013年第5期。

^⑤ 姜国华、岳衡《大股东占用上市公司资金与上市公司股票回报率关系的研究》,《管理世界》2005年第9期。

^⑥ 郑国坚、林东杰、林斌《大股东股权质押、占款与企业价值》,《管理科学学报》2014年第9期。

^⑦ 谢德仁、廖珂《控股股东股权质押与上市公司真实盈余管理》,《会计研究》2018年第8期。

^⑧ 吕晓亮《控股股东股权质押与公司违规》,《山西财经大学学报》2017年第11期。

^⑨ 李晶晶、郭颖文、魏明海《事与愿违:并购业绩承诺为何加剧股价暴跌风险》,《会计研究》2020年第4期。

^⑩ 李晶晶、关月琴、魏明海等《商誉、准则与制度:万亿商誉资产之谜》,《经济管理》2021年第2期。

响的不利资产。商誉大比例减值使得“业绩变脸”在我国资本市场已屡见不鲜,如天神娱乐曾在2018年三季报中预计2018年全年净利润在0至5.1亿元之间,而在正式发布的年报中披露亏损69.78亿元,其中商誉减值金额40.6亿元,占亏损金额的58.18%。上市公司支付较高的商誉成本显著降低了公司业绩^①,增加业绩压力,放大舞弊风险。

本文结合我国资本市场和上市公司现状,选取了关联交易、其他应收款、股权质押、并购和商誉减值比例五个要素加入模型中进行舞弊识别和预测。

综上,本文选取了五个维度的样本数据,具体指标如表1所示。

表1 样本数据类型和指标选取

数据类型	内容
原始财务数据	现金与短期投资、应收账款、存货、流动资产、流动负债、总资产、固定资产、长期负债、折旧与摊销、利息支出、所得税、扣除特殊项目前净收入、流动负债、留存收益、销货成本、净收入、利息收入、短期投资
财务指标	流动比率、现金比率、利息保障倍数、资产负债率、应收账款周转率、存货周转率、总资产周转率、得税率、总资产收益率、营业利润率、净利润现金含量、现金流利息保障倍数、应计项目
治理指标	管理层持股数量、董事人数、前十大股东是否存在关联、董事长与总经理是否兼任、董事与监事及高管年薪总额、委员会设立总数
审计指标	审计费用、审计意见、审计师是否来自国际四大会计师事务所、会计师事务所是否变更
我国资本市场	关联交易占营业收入的比重、其他应收款占总资产的比重、股权质押比例、当期是否发生并购、商誉减值比例
特殊指标	

6. 样本不平衡处理

本文所利用的舞弊数据具有明显的样本不平衡问题,即舞弊样本数量明显少于非舞弊样本。在这种情况下,如果简单利用模型进行分析并以准确率(accuracy)为评价标准,模型若是将所有的样本都视为非舞弊样本,仍然具有很高的准确率,但不具有实际意义,无法为舞弊识别和预测提供帮助。例如,在原始财务数据的样本中,样本不平衡水平为4.64(非舞弊样本数/舞弊样本数),即舞弊样本仅占总样本的17.73%。因此,需要对样本不平衡进行处理。

处理样本不平衡主要采用过采样或者欠采样的方法。过采样的方法主要是重复抽取数据较少的类别以实现降低不平衡水平,但运用重复采样的样本相同,容易出现过拟合问题,降低模型的预测能力。欠采样的方法是在数据量大的类别中随机抽取数据并剔除一部分样本,该方法较为简单,但造成了部分样本数据丢失。本文采用过采样方法处理样本不平衡问题,主要原因有以下两点:第一,过采样的方法能够利用全样本数据,使得模型能够得到更符合资本市场实际情况的训练,减少人为偏差。第二,本文的目标为利用舞弊特征指标进行舞弊识别和预测,重复使用舞弊样本能够使得模型多次利用相关特征指标进行训练,强化特征指标对模型的影响,在测试集中能够更有效地体现出相关特征指标的作用。

7. 样本数据来源与训练集和测试集划分

本文选取我国股票市场2000—2016年所有的上市公司作为研究样本,同时以2017—2020年的公司样本用于样本外检验,测试模型的舞弊识别能力。上市公司舞弊数据和相关指标数据是国泰安数据库(CSMAR)中得到的,其中舞弊数据为根据上交所、深交所和证监会等监管部门发布的

^① 郑海英、刘正阳、冯卫东《并购商誉能提升公司业绩吗?——来自A股上市公司的经验证据》,《会计研究》2014年第3期。

公告中所提及的违规舞弊年度判断得出,若企业某年发生舞弊行为则赋值 1,未发生舞弊则赋值 0。同时,本文将输入样本进行随机分配,其中 60% 的样本生成训练集,40% 样本用于测试集,样本的随机分配一定程度上可以缓解连续舞弊样本造成的结果偏差。^①

四、实证结果及分析

本文首先利用相关数据,评估舞弊预测模型 M-Score、F-Score 以及 C-Score 在中国资本市场背景下的识别和预测能力。之后,以上述三个 Logistic 模型为评价基准,本文进一步利用原始财务数据进行机器学习研究,并依次向模型中加入财务比率、公司治理、审计指标和中国资本市场特殊要素指标,保留能够优化模型结果的指标,逐步输出各模型结果。

从表 2 可知,M-Score 和 F-Score 在我国资本市场中的舞弊识别和预测能力较弱,测试集的召回率仅分别为 0.39 和 0.51,难以达到准确有效识别舞弊的目标。同时,本文利用 C-Score 模型,该模型更适合我国资本市场,结果表明 C-Score 模型能够提高模型的准确率和召回率,分别达到了 0.62 和 0.54。

表 2 M-Score、F-Score 和 C-Score 舞弊识别结果

	M-Score		F-Score		C-Score	
	准确率	召回率	准确率	召回率	准确率	召回率
训练集						
Logistic 模型	0.60	0.37	0.54	0.52	0.62	0.54
测试集						
	0.61	0.39	0.54	0.51	0.62	0.54

注: 召回率为舞弊 [Recall(1)] 的召回率。

原始财务数据的识别和预测结果如表 3 所示。从结果可知,随机森林和 Adaboost-决策树的召回率均为 58%,AUC 分别为 66% 和 65%,高于决策树的相关结果,这符合集成算法优于单一弱分类器的规律。同时,二者的准确率和召回率指标均优于 M-Score、F-Score 和 C-Score,这表明机器学习模型优于现有的 Logistic 模型,机器学习能够更加有效地对舞弊进行识别和预测。但数据结果的准确率和召回率都低于利用美国等国外发达资本市场数据进行的研究^②,这主要是因为我国资本市场正处于快速发展和不断改革的阶段,利用监管部门公布的信息作为判定上市公司舞弊的标准且运用资本市场全样本数据进行分析,难免某些企业存在舞弊行为但未被发现,这种样本数据会对识别和预测结果产生噪音,降低了机器学习的可靠性。与其他方法相比,支持向量机模型的召回率较高,能够正确识别出所有发生舞弊公司中的 80%,对舞弊公司具有较好的识别和预测能力;但准确率只有 41%,将大量非舞弊公司错误预测为发生舞弊的公司,具有较高的预测成本。

本文在后续的研究中逐步加入其他维度数据。如表 4 所示,在加入财务比率数据后,决策树、随机森林和 Adaboost-决策树的召回率分别为 53%、59% 和 59%,与原始财务数据相比,三个模型的召回率分别上升了 2%、1% 和 1%。这表明原始财务数据和财务比率样本的识别预测能力优于只包含原始财务数据的模型,财务比率有助于帮助提升识别效率。支持向量机模型与原始财务数

^① Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

^② Bao , Y. , Ke , B. , Li , B. , Yu , Y. J. , Zhang , J., “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research* , 2020 , 58(1) , pp. 199–235.

据的样本类似,依然具有较高的召回率和较低的准确率。

表3 原始财务数据舞弊识别结果

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.60	0.53	0.64	0.63	0.61	0.69	0.62	0.61	0.67	0.58	0.84	0.65
测试集											
0.64	0.51	0.62	0.64	0.58	0.66	0.63	0.58	0.65	0.41	0.80	0.62

注: 召回率为舞弊 [Recall(1)] 的召回率。

表4 原始财务数据和财务比率舞弊识别结果

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.68	0.71	0.76	0.62	0.64	0.68	0.67	0.72	0.74	0.63	0.83	0.71
测试集											
0.60	0.53	0.60	0.59	0.59	0.63	0.59	0.59	0.62	0.46	0.70	0.59

之后,我们在输入值中继续加入了公司治理指标进行舞弊识别。公司治理能力体现出企业的内部控制水平,内部控制水平越高,公司的财务报告真实性越高。但如表5所示,在加入公司治理指标之后,随机森林和Adaboost-决策树的召回率都出现了不同程度的下降,分别降低了2%和4%。我们认为造成上述现象的原因可能是我国企业的公司治理特征多为制度要求的产物,比如对公司董事会成员人数和独立董事人数的硬性要求。这就造成了董事会结构不是公司自主选择的结果,大量的公司仅是为了满足相关的制度规定而非实际需求。因此,公司治理指标可能难以提升舞弊识别准确性,本文在后续研究中将公司治理指标数据剔除。同时,支持向量机的识别结果也发生了大幅变化,不再具有较高的召回率和较低的准确率。但与随机森林和Adaboost-决策树相比,支持向量机的召回率较低(38%),仍然不具有识别和预测优势。

表5 原始财务数据、财务比率和公司治理指标舞弊识别结果

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.68	0.67	0.75	0.64	0.65	0.71	0.73	0.81	0.83	0.68	0.65	0.77
测试集											
0.62	0.54	0.62	0.62	0.57	0.64	0.60	0.55	0.61	0.60	0.38	0.54

表6展示了在加入审计指标之后机器学习各模型的识别结果。与原始财务数据和财务比率样本相比,决策树、随机森林和Adaboost-决策树三个模型召回率和AUC均有所上升,召回率分别提升4%、1%和2%,AUC分别提升3%、2%和2%,这表明审计指标能够有效提升模型的识别和预测能力。支持向量机的识别结果虽有所提升,但依然较弱。

最后,如表7所示,在加入我国资本市场的特殊要素指标后,决策树的准确率和AUC虽未发生明显变化,但召回率大幅上升4%,识别和预测舞弊公司的效率明显提高。随机森林和Adaboost-决策树的准确率、召回率和AUC均得到明显提升,准确率分别提升1%和3%,召回率分别提升4%和1%,AUC分别提升2%和4%。这表明我国资本市场特殊要素不仅提升了模型的准确性,同时

也降低了舞弊识别的成本。我国资本市场发展起步时间晚,与西方发达国家资本市场存在制度差异。因此,在进行舞弊识别时需要结合我国资本市场特征,这样设计的模型更适合我国的上市公司,有效提升舞弊识别和预测的能力。

表 6 原始财务数据、财务比率和审计指标舞弊识别结果

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.61	0.62	0.67	0.67	0.72	0.73	0.68	0.73	0.76	0.59	0.51	0.59
测试集											
0.60	0.57	0.63	0.61	0.60	0.65	0.61	0.61	0.64	0.62	0.46	0.59

表 7 原始财务数据、财务比率、审计指标和我国资本市场特殊要素舞弊识别结果

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.64	0.67	0.70	0.65	0.67	0.71	0.68	0.70	0.76	0.68	0.58	0.74
测试集											
0.60	0.61	0.63	0.62	0.64	0.67	0.64	0.62	0.68	0.66	0.48	0.63

综上所述,在我国资本市场的环境下,以原始财务数据、财务比率、审计指标和我国资本市场特殊要素组成的上市公司样本集为输入值,采用随机森林模型或 Adaboost-决策树模型进行机器学习识别和预测财务舞弊具有较好的结果。但是,与国内的现有研究相比,本文机器学习预测舞弊的准确率、召回率和 AUC 较弱。我们认为造成这种结果的主要原因是以下两个方面:首先,现有文献可能利用匹配的方式确定样本集^①,舞弊公司与非舞弊公司数量相对平衡,而本文以我国 A 股市场所有上市公司为样本,公司特征差异较大,故降低了识别和预测的结果。其次,部分研究按照审计师出示的审计意见划分舞弊和非舞弊样本^②,而本文依据交易所和证监会发布的公告确定公司是否舞弊,这一定程度上也会造成结果差异。

此外,本文进行了样本外检验,利用 2017—2020 年的我国上市公司为样本,以原始财务数据、财务比率、审计指标和我国资本市场特殊要素作为输入值进行研究,保持模型参数不变,检验模型在样本外的识别和预测能力。具体结果如表 8 所示。

表 8 样本外检验

决策树			随机森林			Adaboost-决策树			支持向量机		
准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC	准确率	召回率	AUC
训练集											
0.67	0.65	0.73	0.69	0.72	0.75	0.74	0.75	0.82	0.88	0.90	0.94
测试集											
0.67	0.62	0.71	0.66	0.65	0.72	0.69	0.63	0.73	0.70	0.41	0.63

从表 8 可知,在利用原始财务数据、财务比率、审计指标和我国资本市场特殊要素的数据集下,

^① 黄志刚、刘佳进、林朝颖《基于机器学习的上市公司财报舞弊识别前沿方法比较研究》,《系统科学与数学》2020 年第 10 期。

^② 高宇《基于贝叶斯网络的企业财务报告舞弊识别研究》,《经营与管理》2018 年第 11 期。

各模型的识别效果均优于 2000—2016 年的样本识别结果, 决策树、随机森林和 Adaboost-决策树的准确率、召回率和 AUC 均出现明显上升。可见, 本文所采用的模型在样本外具有稳健性。同时, 这一识别结果也说明我国资本市场近年来得到了快速的发展, 制度建设更加完善, 监管部门对舞弊的识别越来越有效。这为我们的模型提供了有效数据, 提升了模型的识别能力。

五、研究结论

本文采用机器学习方法研究我国资本市场公司舞弊行为识别和预测。首先, 我们利用 2000—2016 年我国 A 股市场全部上市公司数据, 以上交所、深交所及证监会等监管部门发布的违规舞弊公告形成舞弊样本, 随机生成训练集和测试集。其次, 我们以 Logistic 模型为评价基准(M-Score、F-Score 和 C-Score), 将召回率作为评价各模型及数据有效性的指标, 同时兼顾成本收益原则, 综合考量各模型的召回率和准确率, 并利用 AUC 指标进行辅助评价, 比较各模型的优劣。文章采用过采样的方式降低样本不平衡水平并分别利用决策树、随机森林、Adaboost-决策树和支持向量机模型对样本数据进行分析。最后, 通过比较不同数据和不同模型的有效性, 结果表明, 在原始财务数据基础上加入财务比率、审计指标和我国资本市场特殊要素指标的模型能够得到较优的结果, 而加入公司治理指标会降低模型结果, 这可能是由于我国公司治理特征多为制度规定而非企业自主选择的结果。另外, 本文利用 2017—2020 年我国上市公司的数据进行样本外检验, 结果表明模型在舞弊识别方面具有较强的稳健性。在模型选择方面, 综合考量准确率、召回率和 AUC 的情况下, 随机森林和 Adaboost-决策树具有较好的结果。本文进一步丰富了利用机器学习方法进行公司舞弊识别的研究, 并且在评价指标和数据选择方面做出了新的探索。

A Study on the Fraud of Chinese Listed Companies Based on Machine Learning

YU Li-sheng^{a b}, ZHENG Tian-yu^b, TENG Chuan-hao^b

(a. Center for Accounting Studies, b. School of Management, Xiamen University, Xiamen 361005, Fujian)

Abstract: With the advent of the digital era, the use of machine learning to identify financial fraud has become a hot topic of research. Based on the original financial data, the financial ratio, corporate governance indicators, audit indicators and special indicators of China's capital market, this paper uses the Logistic model (M-Score, F-Score and C-Score) as the evaluation benchmark to conduct machine learning analysis with the decision tree, random forest, Adaboost-decision tree and support vector machine (SVM) models. The sample imbalance is reduced by oversampling, the recall rate is used as the standard to evaluate each model, and accuracy, recall rate and AUC are utilized to judge the model and data. The study has found that the model with the financial ratio, audit indicators and special indicators of China's capital market can obtain better recognition effect, while corporate governance indicators cannot improve the fraud recognition ability of the model. Compared with other models, the random forest model and the Adaboost-decision tree model have better fraud identification effect, with the accuracy rate reaching 62% and 64% and the recall rate hitting 64% and 62%, respectively.

Keywords: digital economy, machine learning, fraud identification, public company

[责任编辑: 叶颖玟]