



机器学习可以有效预警股权质押融资风险吗？——人工介入的必要性和协同效益

焦有章

摘要 上市公司股权质押是一种常见的融资方式，其风险较难控制，许多金融机构深受其害，然而相关风险预警的研究文献很少。本文以 2017—2022 年上市公司控股股东司法冻结事件作为违约样本，运用判别分析、Logistic 回归、BP 神经网络、卷积神经网络、决策树、随机森林、AdaBoost、XGBoost、LightGBM、CatBoost 等方法构建风险预警模型，发现集成学习方法在风险预警方面具有明显优势，其中 XGBoost 正确率最高，达到 91%。尽管机器学习取得了较好的风险预警效果，但是仍存在局限性，在业务实践中还需要与人工介入相结合，金融机构才可以实现良好的经济效益。

关键词 股权质押 股票质押 风险预警 机器学习 XGBoost

一、引言和研究述评

股权质押是一种常见的融资方式，特别是上市公司股票具有权属清晰、流动性好、价格公允的特点，更易被金融机构接受，因此上市公司股权质押融资在资本市场上较为活跃。截至 2023 年末，沪深两市被质押的股票市值为 2.76 万亿元。虽然股权质押能有效满足上市公司股东的融资需求，但同时也面临着较高信用风险。根据同花顺 iFinD 数据，2013—2022 年共有 2 382 家上市公司（非金融行业）的控股股东（含第一大股东）开展过股权质押，其中 529 家发生过股权冻结现象，累计占比 22.21%。如果控股股东所持股份被司法冻结，不仅会导致上市公司股权结构不稳定，还会影响上市公司的正常经营，进而影响投资者的信心，严重的还会影响产业链的

稳定性，对社会经济产生一系列负面影响。股权质押融资风险和司法冻结事件可能成为引发系统性风险的严重隐患，已经引起金融监管部门的高度关注，部分金融机构甚至暂停了股权质押融资业务。因此，如何有效管理股权质押融资风险，已成为当前亟待解决的重要课题。

股权质押融资风险较难控制，且一旦发生，往往需要很长时间来处置，因此事前风险防范措施尤为重要。本文主要研究股权质押融资的风险预警问题，深入分析风险暴露前的征兆和特征，借鉴信用风险预警、财务困境预警、财务舞弊预警、股价崩盘预警等研究思路和研究成果，寻找股权质押融资风险预警的科学方法，帮助金融机构事前识别股权质押融资风险、提升信用风险管理能力，避免计提大额信用减值，从而让金融机构有信心继续开展股

焦有章，东兴证券风险管理部业务总监。本文系个人观点，与所在机构无关。

权质押融资业务，给民营企业提供必要的资金支持，推动实体经济的平稳健康发展。

近年来，股权质押问题引起越来越多学者的关注，尤其是 2017 年后股权质押融资风险集中暴露，相关学术研究迅速升温。经梳理有关研究文献，股权质押易引发委托代理问题（熊礼慧 等，2022）、股价崩盘风险加大（夏常源和贾凡胜，2019）、上市公司会计信息质量变差（谢德仁和廖珂，2018）、控股股东和上市公司行为异常（戴静雯 等，2022）、上市公司信用风险增加（牛煜皓 等，2023）等观点已基本形成共识，并得到了实证检验。然而，对于股权质押融资的事前风险防范，现有研究多集中于案例分析和定性分析，基于实证分析的风险预警研究很少。左扬和丁潘（2019）采用 Logistic 回归建立股权质押风险预警模型，以是否跌至预警线为因变量，研究了市场行情、财务状况、股权结构等变量的预警有效性。刘晶（2020）采用 Logit 回归建立风险预警模型，以是否跌至预警线为因变量，发现退市风险警示、公司属性、大股东持股比例、市盈率、年涨跌幅、流动比率、年振幅、资产负债率、债务资本比等变量具有显著作用。杨柳勇 等（2023）以触及平仓线或司法冻结作为违约样本，运用 XGBoost 预测股权质押风险，发现质押风险高的上市公司在成本费用率、负债率和现金流方面表现较差，XGBoost 模型的召回率达到 74.44%，优于 Logistic 回归模型。尽管这些研究取得了一定进展，但是在样本数据、预警指标方面还需继续完善，在预警效果方面也有待持续提升，在预警模型方面尚缺乏全面系统性的比较。

股权质押融资风险预警主要是预测未来是否违约，可以看作二分类问题。统计

学或机器学习提供了多种分类方法，判别分析、Logistic 回归、BP 神经网络、卷积神经网络、决策树、随机森林、Adaboost、XGBoost、LightGBM、CatBoost 等都可用于构建股权质押融资风险预警模型。本文在前人研究的基础上，全面系统地研究了股权质押融资风险预警问题，旨在寻找有效的风险控制措施，解决当前股权质押融资风险防范的难题。

二、样本数据与预警指标

（一）样本数据

股权质押融资的违约数据不易获取，一些研究使用 Wind 等金融终端中低于预警线的股权质押数据作为违约样本，然而这只是基于股票价格的测算，未考虑融资人补仓和还款的情况，与真实违约情况相比存在较大差异，触及预警线样本的范围要大于真实违约样本的范围。

股权司法冻结是股权质押融资违约后债权人所采取的法律手段，可以视为明确的违约信号。股权司法冻结样本的范围一般小于真实违约样本的范围，因为一部分违约事件发生后股票被平仓卖出，只有当股票存在限售等因素无法卖出时，才会采取司法冻结措施。在实践中，流通股一般能够平仓，平仓后金融机构面临损失的概率较低，但是以流通股进行股权质押融资的情况较少，大部分都是以限售股（或存在减持限制）进行股权质押融资。限售股在出现风险后难以处置，如果涉及多家金融机构，则面临“囚徒困境”，大部分金融机构最终都会采取司法冻结措施，然后进行长达几年时间的诉讼、拍卖，期间股票价格下跌的可能性很大，金融机构面临损失的概率较高。因此，选择司法冻结作为



违约样本，更具有实践意义。

以 2017—2022 年控股股东发生过股权司法冻结的 266 家上市公司（非金融行业）作为违约组；同时随机选择期间存在控股股东股权质押并且未发生过司法冻结的上市公司 266 家作为正常组。正常样本与违约样本的比例为 1:1。然后从中随机各选择 100 家公司作为测试数据集，其余作为训练数据集。考虑到风险预警的时效性要求，违约样本所选指标数据的日期与首次冻结日期至少间隔 1 年，正常组样本的数据日期平均分布于期间各年，以验证金融周期的影响，相关数据来自同花顺 iFinD。

（二）预警指标

在实践中，股权质押融资主要因两种情况触发风险：一是融资人在约定日期无法支付本金或利息；二是质押股票价格下跌，触及平仓线，融资人无法及时补仓或还款。现有文献在研究股权质押融资风险时，采用的预警指标除了控股股东股权质押比例外，还关注上市公司的股价行情、股票估值、财务指标等。这些财务指标以反映上市公司的财务困境为主，而控股股东在股权质押融资发生风险前，可能会出于自身利益操纵上市公司财务报表、侵占上市公司利益，这些行为可以通过财务粉饰或财务舞弊等指标反映出来，因此将反映财务粉饰和财务舞弊的指标纳入研究范围。本文参考多位学者的研究方法和实践中总结的经验，控股股东的股权质押比例、上市公司的现金比率、资产负债率、现金收入比、商誉占总资产比率、其他应收款占总资产比率等财务指标异常，可能会导

致股权质押融资发生风险；同时参考刘云菁等（2022）关于财务舞弊的指标选择、数据处理、模型比较等方法；也参考陈雨露等（2016）的研究方法，设置反映金融周期的相关指标；还借鉴中证指数有限公司的沪深 300 周期行业指数的编制方法，设置周期性行业指标。

在借鉴这些研究成果的基础上，共选取 68 个指标作为备选的预警指标，其中包括财务指标 36 个和非财务指标 32 个。财务指标通常是由资产负债表、利润表和现金流量表等财务报表中的数据生成，这些指标反映了企业的盈利能力、偿债能力、运营能力、盈利质量、资产质量等信息。非财务指标通常是指无法由财务数据来生成的指标，如股权结构、股权价值、股价波动、审计费用、关键审计事项、信用评级等。

在构建风险预警模型时，过多的预警指标可能会增加数据收集和预处理的成本、影响模型的计算效率、增加过拟合风险等，因此有必要对备选预警指标进行筛选。在筛选预警指标时，首先采用 Kolmogorov-Smirnov 检验方法，在 0.05 显著性水平下，所有指标数据均未能通过正态分布检验；然后对连续变量进行 Mann-Whitney 检验，对离散变量进行卡方检验，剔除未通过显著性检验的变量；再执行 Pearson 相关性分析，如果预警指标之间的相关性较高（绝对值大于 0.5），则保留显著性相对较高的指标，剔除显著性相对不高的指标，最终筛选出 36 个指标作为自变量（或特征变量）用于下一步实证分析（表 1）。^① 通过显著性检验筛选预警指标，可以

① 表 1~5，见增强出版，中国知网—《金融市场研究》。

降低数据收集和数据预处理的成本，有助于模型在业务实践中的推广使用。

三、构建风险预警模型

在构建风险预警模型时，使用多种数据分析工具，scikit-learn、pytorch、xgboost、lightgbm、catboost 等 Python 库提供了丰富的机器学习算法，有助于通过参数调优寻找最优解，从而提升风险预警模型的预测效果。

（一）经典方法

判别分析，是 20 世纪 30 年代提出的分类方法，其原理是通过最大化不同类别之间的差异、最小化每个类别内部的差异，来确定一个最优的线性投影方向，使不同类别的样本在投影后能够得到最大程度的分离。采用逐步判别法，筛选出 7 个变量构建模型，在 0.05 显著性水平下，模型通过了显著性检验，且分类效果较好，在测试数据集上正确率为 85%。

Logistic 回归，是 20 世纪中后期由多名学者做出重要贡献的分类方法，随着计算机技术的发展逐渐普及，它以多个指标数据为自变量，将输入的自变量通过 sigmoid 函数映射成 0~1 之间的数据，映射结果可以看作数据样本属于某一类的概率。采用逐步回归法，剔除不显著的变量，最终筛选出 9 个变量构建模型，在 0.05 显著性水平下，模型通过了显著性检验，拟合效果良好，在测试数据集上正确率为 85%。

（二）BP 神经网络和卷积神经网络

BP 神经网络，于 1986 年首次提出，是在早期神经网络算法的基础上发展而来，通常由一个输入层、一个或多个隐藏层、一个输出层组成，在隐藏层和输出层中，每个神经元的加权求和结果会经过一

个激活函数的非线性转换，以更好地捕捉数据模式、提高拟合效果。采用 BP 神经网络构建模型时，输入单元为 36 个（即对应 36 个自变量），设置了 2 个隐藏层，激活函数选择 Sigmoid，输出层的激活函数选择 Softmax。经多次尝试和优化，选出较优的模型，在测试数据集上正确率为 86%。

卷积神经网络（CNN），常用于图像识别、语音识别和自然语言处理等领域，也可以用于处理二维表格数据。CNN 通常由输入层、卷积层、池化层、全连接层组成。输入层和全连接层是整个神经网络的输入和输出，中间的卷积层和池化层使 CNN 能够有效地提取特征，并构建深层的网络结构来解决复杂任务，这是 CNN 区别于普通神经网络的核心结构。基于 Pytorch 搭建 5 个卷积层、5 个池化层、3 个全连接层的 CNN 模型，激活函数选择 ReLU，为提高模型的稳健性，前 2 个全连接层的 Dropout 参数分别设置为 0.07 和 0.03，经过 90 个 epochs 的训练，取得了较好的分类效果，在测试数据集上正确率为 87.5%。

（三）决策树和集成学习方法

决策树，是从数据集中学习出一组规则，并以树状结构表示出来，可用于解决分类问题。根据每个节点的子节点数量，决策树可以分为二叉树、多叉树。采用 CART 二叉树算法构建模型，中间节点的最小样本数设置为 10，叶节点中的最小样本数设置为 5，生成的决策树深度为 2 层，在测试数据集上正确率为 83.5%。

集成学习方法，可以将多个弱学习器组合成一个强学习器，以提高模型的稳定性和准确性。常见的集成学习方法主要有 Bagging、Boosting 等。

随机森林，是一种将 Bagging 思想同



决策树相结合的集成学习方法。它通过随机选择样本和特征变量的子集来训练每棵决策树，然后综合所有树的预测结果，以提升模型的稳健性和准确性。以随机森林构建预警模型时，利用网格搜索法寻找最优参数，对决策树数量、最大决策树深度、最大特征数量、最大样本数量等变量循环搜索，综合考虑正确率、精确率、召回率，选取较优的模型参数，在测试数据集上正确率为 89.5%。

Boosting 是一种基于序列化方法的集成学习方法，它通过对训练集进行加权，每次迭代时选择一个最优的基学习器，并增加那些分类错误的样本的权重，不断进行迭代学习。基于 Boosting 思想的主流算法有 AdaBoost、XGBoost、LightGBM、CatBoost 等，这几种算法各有特点：AdaBoost 采用指数损失函数作为优化目标，对分类错误的惩罚随着错误率的增加而呈指数增长，使模型更加关注那些被错误分类的样本；XGBoost 采用并行计算策略，使用损失函数的二阶泰勒展开近似计算损失函数，引入了正则化项控制模型复杂度，使模型优化更为快速和准确；LightGBM 采用直方图算法，将连续的特征值分桶处理，降低计算的复杂度和内存消耗；CatBoost 具有自适应学习率，能够动态调整每一轮迭代中基学习器的贡献，从而提高模型的精度。采用决策树为基学习器，利用网格搜索法寻找最优参数，AdaBoost、XGBoost、LightGBM、CatBoost 在测试集上的正确率分别为 88.5%、91%、88.5%、89%。

（四）其他方法

在构建预警模型过程中，还尝试了其他分类方法，例如，支持向量机（SVM）

正确率为 79%、朴素贝叶斯（Naive Bayes）的正确率为 74.5%、K-近邻（K-Nearest Neighbors）的正确率为 64%。这三种方法在自动选择变量方面存在局限，如果借助其他技术预先筛选变量或降维，它们的正确率可以提高到 84.5%、83%、73%。由于所用的指标数据无法与其他分类方法保持一致，限制了可比性，且分类效果不具有优势，因此未对这三个模型进行详细比较分析。

四、模型比较与相关建议

（一）模型预测效果比较

基于相同的样本数据，计算了 10 种模型预测效果的评价指标（表 2）。从评价指标看，BP 神经网络和卷积神经网络的预测效果优于 Logistic 回归和判别分析，然而神经网络模型的预测效果具有一定波动性。在测试卷积神经网络模型时，采用不同的随机种子可能得到不同的预测效果；即使采用相同的随机种子，在不同计算机上也可能得到不同的预测效果；当数据集的变量顺序发生变化时，也可能得到不同的预测效果。在实践应用中，我们应谨慎对待这种波动性，可能需要多次测试以获取稳定的预测结果，并考虑保持生产环境与测试环境的一致性，保持样本数据结构的一致性。

决策树在正确率、精确率、召回率、F1-Score、AUC 等指标上表现最差，但是以决策树作为基学习器的 XGBoost 和随机森林等集成学习方法却表现很好，XGBoost 的正确率和 F1-Score 最高，随机森林的召回率最高，CatBoost 的 AUC 最高。因为单个决策树容易受到噪声的影响，预测效果往往不佳，而集成学习方法可以对

多个决策树的分类结果进行综合分析以改善模型预测效果。

（二）模型稳健性比较

通过比较模型在训练数据集和测试数据集上的表现，可以评估模型的稳健性。如果模型在测试数据集与训练数据集上的表现接近，表明模型具有良好的稳健性。从相关预警模型在训练数据集和测试数据集的正确率差异看（表3），Logistic 回归、判别分析和决策树的稳健性较好；集成学习方法可能存在一定程度的过拟合问题，导致它们在测试数据集上的表现逊色于训练数据集。在集成学习方法中，XGBoost 显示出相对较好的稳健性，它在训练数据集和测试数据集上的正确率差异相对较小。

（三）重要预警指标分析

重要预警指标是指预警模型中发挥重要作用的变量。从各模型所采用的主要变量看，既存在共性，也存在差异。不同模型评估变量重要性时，采用了不同方法，例如，标准化典则判别系数（判别分析法）、瓦尔德系数（Logistic 回归）、权重（神经网络）、信息增益（决策树）、Shapley 值（CNN）等。在综合评估变量重要性时，每个模型最多取前 10 个变量。根据统计结果（表4），尽管不同预警模型基于不同的技术方法，并且采取了不同算法对自变量的重要性进行评估，但它们都会重视控股股东股权质押比例、预付款项占营业收入比率等变量的作用，尤其是控股股东股权质押比例，在 10 个模型中均被认为是最重要的变量。

从实证分析结果看，金融机构在评估股权质押融资风险时，应重视以下五方面因素。

①应重视融资人的股权质押比例。控

股股东股权质押比例是最重要的预警指标，当股权质押比例较低时，大部分金融机构愿意接受其再融资，融资人可以通过滚动融资为债务续期；当股权质押比例较高时，再融资时可选的金融机构较少，再融资能力受到限制。无限售股质押比例也是重要的预警指标，因为无限售条件的股票易于卖出，常被认为是优质的质押物，金融机构更愿意接受。如果融资人的优质资产质押比例较高，是偿债能力弱化的标志。

②应重视上市公司的财务指标异常。上市公司的预付款项、其他应收款等指标异常，可能是控股股东已经甘愿冒着违法违规的风险侵占上市公司利益，通常预示着严重的债务压力。上市公司的总资产周转率、债务融资成本、ROE 等指标，反映了上市公司的经营情况，当这些指标出现问题时，股票价值可能会受到影响，更容易触及平仓线。

③应重视上市公司股票的价值。最近一年股票涨幅、市盈率等指标也是有效的预警信号。股票价格长期下跌、市盈率的调整，可能会降低融资人持有股票的价值，从而影响融资人的偿债能力。

④应重视第三方对上市公司的评价。审计费用、研究机构关注数量、评级下调或负面等指标，分别代表了审计机构、研究机构、信用评级机构对上市公司的评价。提高审计收费可能是审计机构为避免受上市公司影响所采取的措施，以补偿其面临的潜在风险。研究机构关注数量较少，可能表示上市公司经营或股票价值并未受到广泛认可。信用评级下调或负面，是信用评级机构对上市公司偿债能力的综合评价。

⑤应重视金融周期的影响。金融周期的波动对企业的信用状况和违约率有显著



影响。当金融周期处于高涨期（顶部）时，中央银行一般会采取紧缩的货币政策，随后信用开始收缩，企业的违约率上升。当金融周期处于衰退期（底部）时，一般会采取宽松的货币政策，随后信用开始扩张，企业的违约率下降。

（四）模型局限性及解决措施

尽管机器学习模型在风险预警方面取得了良好效果，但是将它们直接应用于实际业务场景时，仍存在一些需要解决的问题。由于股权质押融资的利差较低，大约在 2%~3% 左右，而模型预测的最高正确率为 91%，如果仅依赖于机器学习技术构建的风险预警模型，而不进行人工干预，金融机构显然难以取得良好的经济效益。

对于不同的数据样本，机器学习模型的分类效果可能存在差异，尤其是对分类边缘附近的样本，模型的区分度往往较差，例如，以分类边缘附近的 10% 样本作为研究对象，XGBoost 模型的正确率会由原来的 91% 下降到 80%。如果金融机构单纯依赖机器学习模型，可能会引发“逆向选择”问题。在短期绩效考核的驱动下，业务开展时可能更倾向于选择那些勉强归属正常组的业务，这类业务的拓展难度相对较小，但是其风险高于正常组的平均水平，真实风险会高于模型预测的结果，可能会给金融机构带来很大风险隐患。这也是金融机构容易被前沿技术误导的关键风险点，需要制定和执行正确的风险策略进行干预。

样本数据的完整性、准确性、代表性都会对预测结果产生较大影响。如果样本数据难以获取，未进入模型，可能会产生有偏的预测结果，例如，通过公开渠道无法获取融资人的对外担保、融资利率等信息，可能影响模型的预测能力。不准确的

样本数据会产生错误的预测结果。还有一些变量的异常值出现概率很低，对分类结果具有较大影响，但是模型不一定能捕捉到这种低概率的异常情况。

针对上述模型缺陷，在人工介入时可以从以下几方面采取措施。

①从模型整体角度，根据模型给出的分类概率对业务进行排序，设置合适的业务准入标准，舍弃一些处于违约边缘的业务，有助于提高金融机构的资产质量，但是仍不能直接将模型用于业务实践。例如，在 XGBoost 模型预测不会违约的样本中，如果分别舍弃违约概率最高的 10%、25%、50%，剩余样本构建的资产组合中违约率分别为 8.70%、8%、10%，对业务实践来说，违约率仍然较高。

②从预警指标角度，在单个预警指标上设置更高的准入标准，弥补模型可能忽略的某些小概率规律或异常值。“事出反常必有妖”，舍弃一些存在指标异常的业务，有助于降低股权质押融资的违约率。例如，在 XGBoost 预测不会违约的样本中，将控股股东股权质押比例、预付款项占营业收入比率、营业收入同比变化、资产负债率、EBIT 利息保障倍数、Z-SCORE、审计费用等指标的 10% 分位数设置为准入标准，舍弃了约 47% 的样本，用剩余样本构建资产组合，违约率约为 2%，接近于银行业的不良贷款率，金融机构可以实现良好的经济效益。

③从样本外角度，由于数据获取难度较大，一些包含重要风险信息的变量未被包括在模型中，例如，控股股东的负债规模和对外担保规模、历史质押记录中的债权人、股票质押率、历史融资利率等信息。在评估股权质押融资风险时，还应重点关

注这些未纳入模型的变量，通过人工进行分析和评估，可以更全面地评估股权质押融资风险。

五、进一步观察

在实际业务场景中，正常业务与违约业务的比例并不均衡。为了进一步验证模型预测效果、重要预警指标、模型局限性及解决措施的普适性，本文将正常样本数量扩充为 1 064 个，违约样本数量不变，正常样本与违约样本的比例为 4:1，接近于实际业务场景的违约率，按照 7:3 比例划分为训练数据集和测试数据集。为减少类别不平衡带来的偏差，采用 SMOTE 过采样技术处理训练数据集，然后计算 10 种模型分类预测的评价指标（表 5）。

与之前样本的预测效果相比，10 种模型在测试数据集上的正确率平均提高了 0.76%，召回率平均降低了 14.43%，AUC 平均降低了 3.19%。集成学习模型仍然表现良好，XGBoost 的正确率、精确率和 AUC 最高。控股股东股权质押比例、无限售股质押比例、预付款项占营业收入比率、债务融资成本等指标重要性仍然很高，然而 SMOTE 增加了少数类样本的多样性，有助于模型学习到更多的特征和模式，例如，审计意见、研究机构关注数量等指标的重要性提高。

机器学习模型难以应对“逆向选择”的局限性仍然存在，例如，对于分类边缘附近的 10% 样本，XGBoost 的正确率由 91.54% 下降到 85.42%。但是，SMOTE 过采样技术解决了数据分布不均的问题，在面临“逆向选择”问题时，模型性能还是有了较大改善。如果从预警指标角度筛选，将控股股东股权质押比例、预付款项占营

业收入比率、研究机构关注数量、关键审计事项等指标的 10% 分位数作为业务准入标准，舍弃约 33% 的样本，剩余样本构建的资产组合可实现违约率低于 2% 的预期目标。由此可见，异常的预警指标确实是识别股权质押融资风险的有效信号。

使用不同的数据样本，进一步检验了机器学习模型的风险预警效果，也检验了人工干预弥补模型局限的必要性和合理性，机器学习和人工介入的协同应用可以让金融机构实现良好的经济效益。

六、结论

上市公司的股权质押融资风险较高，许多金融机构深受其害，然而相关风险预警的研究文献很少。过往文献以 Wind 等金融终端测算的触及预警线数据作为违约样本，但是与真实样本存在一定偏差，本文以司法冻结作为违约样本在业务实践上更具有代表性，在采用多种机器学习方法构建风险预警模型后，发现集成学习方法在风险预警方面具有明显优势，其中 XGBoost 正确率最高，达到 91%。综合分析多个模型的预测结果后发现，最重要的预警指标是控股股东股权质押比例。

尽管机器学习技术在风险预警方面取得了良好效果，但仍存在一些局限性。如果单纯依赖模型，股权质押融资业务的利差尚不足以覆盖风险损失。模型对违约边缘样本的预测效果相对较差，难以解决“逆向选择”问题。一些获取难度较大的变量数据未进入模型，影响了预警模型的准确性和直接应用。为了克服这些局限性，需要将机器学习技术与人工介入相结合，以重要预警指标的 10% 分位数作为准入标准，舍弃一些风险较高的业务后，构建资



产组合的违约率约为 2%，接近于银行业的不良贷款率，金融机构可以实现良好的经济效益。

金融机构应积极跟踪并深入了解机器学习等前沿技术的优势，以充分发挥这些技术在风险管理中的作用。同时，金融机构必须正视机器学习技术的局限性，避免

仅仅为了追求技术的新颖性而盲目采用。在面对复杂多变的金融市场时，适当的人工介入至关重要。结合机器学习的优势和专业人员的经验判断，金融机构可以实现更好的协同效益。^[N]

学术编辑：韦燕春

参考文献

- [1] 熊礼慧,董希淼,卢泉.控股股东股权质押是否会侵占中小投资者的利益——基于A股上市公司的实证[J].金融监管研究,2022(2):65-80.
- [2] 夏常源,贾凡胜.控股股东股权质押与股价崩盘:“实际伤害”还是“情绪宣泄”[J].南开管理评论,2019,22(5):165-177.
- [3] 谢德仁,廖珂.控股股东股权质押与上市公司真实盈余管理[J].会计研究,2018(8):21-27.
- [4] 戴静雯,许荣,冯迪.控股股东股权质押会导致公司违规增加吗——基于公司治理视角的考察[J].金融监管研究,2022(7):22-40.
- [5] 牛煜皓,卢闯,鲍睿,等.控股股东股权质押、质押压力与公司违约风险[J].中央财经大学学报,2023(5):63-73.
- [6] 刘云菁,伍彬,张敏.上市公司财务舞弊识别模型设计及其应用研究——基于新兴机器学习算法[J].数量经济技术经济研究,2022,39(7):152-175.
- [7] 左扬,丁潘.上市公司股权质押风险预警研究[J].科技创业月刊,2019,32(10):37-39.
- [8] 刘晶.基于Logit模型的A股上市公司股权质押融资风险预警研究[D].北京:中国科学院大学,2020.
- [9] 杨柳勇,岑盛楠,贾少卿.上市公司股权质押风险特征识别与测度[J].金融市场研究,2023(12):72-80.
- [10] 焦有章.股票质押项目评审与风险防范[M].天津:天津人民出版社,2021.
- [11] 陈雨露,马勇,阮卓阳.金融周期和金融波动如何影响经济增长与金融稳定?[J].金融研究,2016(2):1-22.

Can Machine Learning Effectively Predict Equity Pledge Financing Risks? —The Necessity of Human Intervention and Synergistic Benefits

JIAO Youzhang

(Risk Management Department,Dongxing Securities)

Abstract Equity pledges of listed companies are a common financing method, but one drawback is that risk control is difficult. Many financial institutions have incurred financial losses. This paper takes as default samples the judicial freeze events of controlling shareholders of listed companies over the period of 2017 to 2022. It constructs risk-warning models using discriminant analysis, logistic regression, BP neural network, convolutional neural network, decision tree, random forest, AdaBoost, XGBoost, LightGBM, CatBoost, and other methods. The study finds that Ensemble Learning methods have obvious advantages in risk warning, with XGBoost having the highest accuracy at 91%. Although machine learning has shown promising results in risk prediction, there are still limitations. In practical business operations, a combination of machine learning and human intervention is necessary for financial institutions to achieve optimal economic benefits.

Keywords Equity Pledge, Stock Pledge, Risk Warning, Machine Learning, XGBoost

JEL Classification G32 C51 C45