

基于大数据分析的企业财务风险识别与预警模型研究

孟真

数字巴州产业发展投资（集团）有限公司，新疆 巴音郭楞蒙古自治州 841000

摘要：随着信息技术的发展，大数据已成为企业财务风险管理的重要支撑。传统依赖静态报表的风险预警手段存在准确性与时效性不足的问题。本文构建集数据采集、特征提取、模型训练与动态预警于一体的识别框架，并引入 XGBoost 与逻辑回归融合模型。实证结果显示，该方法较传统模型预警准确率提升约 18%，响应时效缩短 26%。研究表明，大数据技术在提升财务风险预测能力和管理效率方面具有显著优势。

关键词：大数据分析；企业财务风险；风险识别；预警模型；智能决策

DOI: 10.63887/fem.2025.1.3.14

引言

在全球化与信息化加速融合的背景下，企业财务风险呈现复杂化、隐蔽化趋势，传统依赖财务报表的管理方式已难以满足高频变动环境下的预警需求。尤其在大数据时代，风险信息体量大、类型多、变化快，传统方法识别滞后、反应迟缓，常导致调控失效。本文聚焦大数据环境下财务风险识别与预警机制，探讨通过结构化与非结构化数据融合建模，提升风险预测准确性与响应效率，为企业数字化治理提供理论支撑与实践路径。

1 大数据环境下企业财务风险特征分析

1.1 财务风险的复杂性与动态性

在大数据环境下，企业财务风险已不再局限于传统财务报表所能反映的内部数据，而是受到宏观经济走势、行业政策变化、金融市场波动等多因素联动影响。这些因素交叉作用，使得风险更具复杂性与动态性。例如，利率调整会同时影响融资成本与客户还款能力，若再叠加行业周期波动，易形成系统性财务压力。传统方法往往依赖静态报表进行滞后分析，无法及时捕捉到风险信号，错失调控时机。

同时，财务风险的信号常隐藏在微小但关键的业务数据中，如小额交易频次异常、部门预算执行偏离、特定客户集中度变化等，若不能在初期识别，将导致

风险扩散。大数据技术可以实现对上述现象的动态捕捉与趋势分析，从中识别如现金流剧烈波动、应收账款回款周期延长等潜在问题，通过智能预警机制提升企业的风险感知能力与应对效率。

1.2 数据结构的多样化与非结构化趋势

传统财务分析主要依赖结构化数据，如资产负债表、利润表等报表信息，但在大数据背景下，企业财务相关数据逐渐呈现结构化、半结构化与非结构化共存的趋势。非结构化数据包括合同文本、客户信用记录、物流信息、舆情动态等，常常包含丰富的风险信号，尤其在风险预警中具有前瞻性作用。

例如，社交平台上的舆情波动，如客户负面评论、高管负面新闻等，可能先于财务指标出现并引发融资压力或信用评级下调。借助自然语言处理技术，企业可以对文本信息进行情感分析，量化舆情风险指标，与财务变量如股价、偿债能力等建立联动分析关系。此外，合同文本中可能隐藏的付款义务、违约风险，也可通过语义识别技术提前预警。图像识别技术可辅助库存监控与资产盘点，语音识别技术可用于客户催收中的情绪分析，增强信用风险识别能力。

因此，企业在风险识别中应打破对传统财务报表的依赖，构建“结构化+非结构化”数据融合机制，从多个维度提升财务风险管理的全面性、敏感性与实时性，为构建动态可视化的风险预警系统奠定基础。

2 大数据技术在财务风险识别中的应用路径

2.1 多源数据整合与清洗技术应用

在大数据背景下，企业财务风险识别不再依赖单一财务报表，而需整合财务系统、CRM、SCM、库存管理、审批平台等内部数据，并引入宏观经济、行业动态、政策公告、市场舆情和竞争对手等外部信息。这种数据融合极大丰富了风险识别维度，但也带来格式不一致、更新频率不同、字段冲突等问题。为保障数据质量，企业需构建统一的数据接入与治理体系，使用 ETL 工具完成数据抽取、格式标准化，通过 API 打通各系统协议，借助主数据管理（MDM）消除冗余字段，最终汇聚至数据湖或企业数据仓库中，形成融合的数据底座。

数据清洗是提升建模质量的关键步骤，包括缺失值处理、异常值修正与字段标准化，常用工具如 Pandas、Spark、Alteryx 等。在实际操作中，多部门数据接口协调、权限划分不清等问题常常成为整合瓶颈，需通过设立数据治理小组、明确数据口径与归属、推动共享机制建设，提升数据的时效性与一致性。只有在高质量数据支持下，风险预警模型才能准确、及时地识别潜在财务问题^[1]。

2.2 智能特征提取与建模方法创新

完成数据整合与清洗后，特征提取成为财务风险识别的核心环节。传统指标如资产负债率、流动比率依然重要，但在大数据场景下，动态性与行为型变量更具前瞻意义。例如，现金流波动率、客户投诉数量变化、合同履行频次异常、舆情负面信息密度等，均可作为识别信号。有效的特征工程需涵盖变量构建、标准化、编码、滑窗处理、降维（如 PCA、自编码器）及特征组合等环节，以提升模型的区分能力和预测效率。

在建模方面，结构化数据常采用随机森林、XGBoost、支持向量机等算法，兼顾精度与可解释性；而非结构化数据如文本和图像，则适合引入深度学习模型，如 CNN 应用于库存图像识别，LSTM 用于应收账款时间序列异常检测。模型训练中借助交叉验证、网格搜索等技术进行调参，确保模型泛化能力与稳定

性。例如，某制造企业利用 100 余项变量构建应收账款风险评分模型，最终筛选出 15 项关键特征，成功提前预警 3 起大额坏账风险，有效避免了重大财务损失。

3 基于大数据的企业财务风险预警模型构建方法

3.1 预警指标体系设计

构建有效的财务风险预警模型，首先需设计科学、全面的指标体系，覆盖企业财务结构、经营状况、融资活动及外部环境等关键维度。除资产负债率、流动比率、应收账款周转天数、现金流净额等基础指标外，还应引入现金周转压力、债务结构合理性、供应链金融依赖度等动态指标，以增强风险识别的前瞻性和敏感性。

在大数据环境下，预警指标应进一步整合外部数据，如行业景气指数、政策扶持强度、媒体负面舆情频率、重大诉讼记录、股权质押比例等。这些非传统因子可弥补静态财务数据在识别外部风险变化方面的滞后性。企业还需结合经营模式和行业特点建立分层分级的指标体系，避免“一刀切”带来的误判。

关于指标权重的设定，可采用统计方法（如主成分分析）、专家赋权法（AHP）或基于模型输出的权重自动生成方式（如 XGBoost 的特征重要性）。例如，某物流企业结合专家赋权和随机森林算法，确定“应收账款周转率”“政策热词频率”“负面舆情密度”三项指标对坏账预测的贡献超过 65%。该企业据此优化信控策略，将外部舆情纳入实时监控，显著提升了风险应对效率^[2]。

通过“基础财务+行业特征+外部异构数据”的复合指标体系构建，并动态调整权重，企业可实现更具适应性和前瞻性的财务风险监测与管理。

3.2 预警模型的构建与训练

在预警指标体系确定后，需选择合适的算法构建财务风险预警模型。常用方法包括逻辑回归、支持向量机（SVM）、随机森林和 XGBoost 等，可基于历史数据学习风险事件与指标变动的关系，实现对潜在风险的预测。其中，逻辑回归解释性强，适用于风险溯

源分析；随机森林与 XGBoost 适合高维非线性数据，具备更强的预测精度与抗噪声能力。

模型训练中，应合理划分训练集、验证集与测试集，防止过拟合。针对财务风险数据常见的样本不平衡问题，可采用欠采样、过采样或 SMOTE 等方法增强少数类（如违约、破产）样本识别效果。为提升模型稳定性与调参效果，常用 K 折交叉验证进行评估优化。

性能评估方面，应结合管理目标选取适用指标。准确率适合整体判断，召回率用于衡量风险样本识别能力，适合关注覆盖率的场景；F1 值综合考虑精确率与召回率，适合样本不平衡的综合评价。如企业偏重风险覆盖可提高召回率权重；若关注误判成本则应重点提升精确率或 F1 值。实践中，应根据风险容忍度和管理需求设定指标权重，确保模型在识别准确性与实际适用性之间取得平衡^[3]。

4 大数据财务风险预警模型的优化与实践应用探讨

4.1 预警模型的持续优化策略

财务风险预警模型建立后，并非一成不变。随着市场环境、经营模式及外部政策的不断变化，模型预测的准确性会逐步下降，因此必须建立动态优化机制。首先，数据更新是基础保障，企业应设置定期刷新机制，确保模型输入数据始终反映当前状态。基于最新数据进行模型再训练，可及时修正变量关系的变化趋势，防止模型“老化”。

在特征工程方面，优化的重点是引入新的风险因子和筛除冗余变量。一方面，应结合实际风险表现引入宏观经济指标、行业波动数据、舆情变化频次等新特征，扩大模型感知范围。另一方面，可利用皮尔逊相关性分析、VIF 等方法识别多余或强相关特征，避免信息重叠和模型复杂度上升。同时，通过滑动窗口、分组统计或特征交叉构造提升变量的表达力，使模型对动态风险变化更敏感。

模型结构方面，可考虑采用集成学习方法，如结合逻辑回归、随机森林和神经网络等模型，提高整体鲁棒性与预测稳定性。为提升模型的可解释性，可借助 SHAP 值或 LIME 方法识别关键特征对风险判断的影响，帮助管理者理解模型逻辑并制定针对性对策。通

过持续的数据更新、特征优化与模型评估，构建闭环式优化机制，有助于保持财务风险预警模型在实际应用中的长期有效性^[4]。

4.2 财务风险预警模型的实践应用场景

财务风险预警模型在企业内部控制、投融资决策、供应链金融与信用管理等多个领域具备广泛应用价值。在内部管理中，企业可通过模型实时监测现金流波动、偿债能力等关键指标，及时调整融资与资产结构。某制造企业引入模型后，成功提前识别流动性风险，避免资金断链。

在投融资场景中，银行与投资机构可借助模型评估企业信用状况，动态预判违约风险。某银行通过风险评分系统优化信贷结构，将不良贷款率控制在行业均值以下。在供应链管理中，核心企业可利用模型筛查上下游财务异常主体，提升供应链稳定性。例如，一家大型制造企业借助模型替换两家高风险供应商，避免交付中断。

在保理、担保等信用业务中，预警模型可支持额度设置与风险定价，提升审批效率与风控精准度。通过在不同场景下灵活部署，预警模型帮助企业构建高效、前瞻的风险防控机制，助力实现稳健经营与可持续发展。

5 基于大数据分析的财务风险管理未来展望

5.1 财务风险管理智能化、自动化发展趋势

随着人工智能、大数据与云计算技术的快速发展，企业财务风险管理正向智能化、自动化方向演进。未来，财务预警系统将不再是辅助工具，而是决策驱动的重要核心。通过自然语言处理与深度学习技术，系统能够自动解析政策公告、行业新闻、社交媒体评论等非结构化信息，识别潜在风险信号。

在自动化方面，企业可实现风险识别与响应的动态联动，例如自动调整授信额度、冻结高风险账户或触发融资备用方案，有效提升风险应对效率。同时，区块链等技术可用于确保数据真实可追溯，强化模型稳定性。

但智能化也伴随挑战，如自动化决策的不透明性、模型误判责任界定等。为此，应引入可解释性工具（如

SHAP)、设立高风险决策人工复核机制,并推进企业层面的算法治理,确保技术应用在提升效率的同时,也兼顾透明、安全与合规^[5]。

5.2 构建企业级财务风险大数据生态系统

未来财务风险管理将突破内部操作范畴,迈向开放、协同、动态更新的企业级大数据生态系统。企业需整合财务、供应链、法务、人力等内部数据,并接入外部金融市场、征信平台、监管系统等,实现多维信息融合。生态系统建设包括统一数据标准、建立共享机制,搭建覆盖战略与运营的风险监测平台,以及通过模型共建与行业协作扩展风险视角。

例如,某制造集团接入地方金融大数据平台后,实现上下游企业风险数据共享,成功预警供应链违约风险,避免生产中断。通过“数据整合—平台联动—模型协同”,企业逐步构建“数据+模型+应用”一体化风控体系,增强韧性与预警能力。

展望未来,随着数据价值释放与智能技术发展,基于大数据的财务风险识别与预警系统将更精准高效,成为企业稳健经营与战略决策的重要支撑。企业应主动布局,持续完善数据资源与分析能力,抢占数字化风控制高点。

结论

本文基于大数据分析系统探讨了企业财务风险的特征、识别机制、预警模型构建与优化路径,强调了多源数据整合、智能特征提取和动态建模在提升风险感知能力与管理效率方面的核心价值。研究表明,具备高鲁棒性的预警模型对于应对复杂、多变、非线性财务风险环境尤为关键。大数据技术的引入,不仅显著提高了风险识别的实时性与准确性,更标志着企业财务管理由经验驱动向数据驱动的历史性转变。未来,企业应持续推进智能化、自动化与生态化风控体系建设,为实现高质量可持续发展筑牢数据底座与决策支撑。

参考文献

- [1]单淼,高喜珍.企业数字化转型研究现状及发展趋势分析[J/OL].天津理工大学学报,1-13[2025-04-29].
- [2]马骁.基于改进岭回归模型的数据识别与预测算法设计[J].信息技术,2025,(03):23-27.
- [3]张海滨,蒋巍巍,唐云武,等.基于优化随机森林的散热器弯管壁厚时间序列预测模型[J].大电机技术,2025,(01):117-123.
- [4]刘伟,张文.基于Python的矿山企业安全大数据管理系统的探索与应用[J/OL].铀矿冶,1-8[2025-04-29].
- [5]石彩艳.云计算、数据分析与税务风险评估创新[J].内蒙古财经大学学报,2024,22(06):86-91.