

人工智能在财务舞弊识别中的模型构建与应用

李 铭 姜国鹏

黄河水利委员会河南水文水资源局 河南 郑州 450000

摘要：财政资金监管趋严下，事业单位财务行为规范性关乎公共资源安全效能。但部分事业单位因内控弱、监督滞后，存在虚列支出等隐蔽舞弊行为。传统人工审核与静态规则监督方式，面对多源动态财务数据，难高效精准预警风险。近年，AI技术凭借模式识别与自适应学习能力，为事业单位财务舞弊识别提供新路径。本文结合事业单位特点，梳理其财务舞弊典型表现与成因，探讨适用于事业单位的AI模型构建流程，分析逻辑回归等技术在不同场景中的适用性与优势。同时指出，AI在事业单位应用面临数据孤岛等现实挑战，并对构建“智能+制度”融合的财务风险防控体系提出展望，为提升事业单位财务治理现代化水平提供理论与实践参考。

关键词：人工智能；财务舞弊；机器学习；深度学习；模型构建；异常检测

引言

事业单位承担公共服务职能，资金具公共性与政策导向性，财务舞弊行为如虚报经费、挪用专项资金等，会造成财政资源浪费，损害政府公信力与公共服务质量。近年来审计署披露多起违规案例，凸显加强事业单位财务风险防控的紧迫性。传统监督手段依赖内部审计、财政检查和手工凭证审核，存在“事后发现”“抽样有限”“规则僵化”等缺陷，难以捕捉隐蔽舞弊模式。在此背景下，人工智能技术凭借对高维、非线性、时序数据的处理能力，为构建主动式、智能化财务风险识别体系提供可能，在预算执行监控等高频场景中可实时感知与动态评估异常行为。因此，研究AI驱动的财务舞弊识别模型意义重大。

1 财务舞弊的特征与成因分析

1.1 舞弊三角理论

尽管事业单位不以盈利为目的，但其仍面临绩效考核、项目结题、预算执行率等多重压力。例如，部分单位为完成“预算执行进度”指标，年底突击花钱；科研团队为维持课题延续性，虚列劳务费或测试费。这些构成了舞弊的“压力源”。“机会”则源于内控制度执行不严、岗位职责不清、信息系统割裂等问题。如财务、采购、资产管理数据未打通，导致同一设备重复报销；或审批权限过于集中，缺乏有效制衡。“借口”常表现为“为了单位发展”“项目急需”“惯例操作”等自我合理化说辞，削弱道德约束。三者叠加，形成事业单位特有的舞弊土壤。

1.2 财务舞弊的典型表现

在事业单位实践中，财务舞弊常通过扭曲财务与业务数据留下可识别的异常痕迹。例如，部分单位为提高

预算执行进度，在年末集中虚增会议费、咨询费或办公用品采购，导致相关科目支出激增，但缺乏真实业务支撑；又如，科研经费被用于发放人员补贴或报销私人消费，表现为劳务费占比畸高、发票内容模糊或供应商高度集中^[1]。此外，固定资产长期不盘点、报废不销账，造成账实严重不符；或通过关联方以虚高价格采购设备，实现资金转移。这些行为虽不直接体现在资产负债表或利润表上，却在报销明细、支付流水、合同文本等操作层数据中形成多维度、跨期间、结构性的异常模式，为人工智能模型提供了丰富的判别依据。

2 人工智能模型构建流程

2.1 数据收集与标注

构建适用于事业单位的AI舞弊识别模型，首先依赖于高质量、多源融合的数据基础。数据不仅包括财政预算管理中的项目预算与执行记录、单位内部财务核算系统的凭证与科目明细，还应涵盖电子报销平台的单据信息（含发票、审批流、事由说明）、资产与采购系统的业务日志，以及外部数据如税务发票查验结果、政府采购中标公告等。在数据标注方面，通常将经审计机关认定、纪检监察立案或财政绩效评价中明确指出存在违规使用资金的事项标记为正样本，其余视为正常样本。然而，由于事业单位舞弊事件本身属于低频、隐蔽行为，正样本极度稀疏，导致训练集严重不平衡，模型易偏向多数类而忽略真实风险。因此，必须在后续环节引入合成少数类过采样（SMOTE）或代价敏感学习等策略，以提升对舞弊案例的敏感度。

2.2 数据预处理

原始财务与业务数据普遍存在缺失、异常、口径不一等问题，需经过系统化预处理以提升模型输入质量。针

对事业单位会计科目编码不统一、财政专户与基本户资金混用等情况，需先进行标准化映射与账户分离。对于缺失值，可根据变量特性采用中位数填充或基于时间序列的插值法；对极端异常值，则通过四分位距（IQR）或Z-score方法识别并修正，避免干扰模型训练。此外，不同财务指标量纲差异显著（如总支出与单笔报销金额），需进行归一化或标准化处理^[2]。尤为重要的是，针对样本不平衡问题，除采用过采样技术外，还可设计加权损失函数，在模型训练阶段赋予舞弊样本更高的误分类惩罚权重，从而引导算法更关注高风险行为的识别。

2.3 特征工程

特征工程是连接原始数据与模型性能的关键桥梁。在事业单位场景中，特征构建需兼顾财务规范与业务逻辑。除基础比率（如预算执行率、三公经费占比）外，应重点开发契合公共部门特点的复合指标，例如“报销频次异常指数”（同一人员短期内高频小额报销）、“供应商集中度”（单一供应商占同类支出比重过高）、“跨科目协同异常”（会议费与差旅费同步激增但无对应活动记录）等。同时，静态指标往往不足以揭示风险，需引入趋势性特征，如季度末支出突增幅度、年度内报销节奏偏离历史均值等。近年来，随着自然语言处理技术的发展，从报销事由、项目申报书等文本中提取语义特征也成为重要方向。例如，通过词嵌入技术量化表述的模糊性（如频繁使用“其他”“相关费用”）或逻辑矛盾（如申报成果与实际支出不匹配），这些非结构化信息往往能揭示数字背后隐藏的动机与风险。

2.4 模型选择与训练

在模型选择上，需综合考虑数据特性、业务需求与制度环境。逻辑回归和支持向量机结构简单、决策透明，适合对可解释性要求较高的内控与审计场景；而集成学习模型如XGBoost和LightGBM通过组合多个弱学习器，在处理高维非线性数据时表现出卓越性能，已成为当前主流选择。深度学习模型如LSTM可有效捕捉预算执行或报销行为的时间依赖性，但其“黑箱”特性在强调责任追溯的公共部门应用受限。更为前沿的图神经网络（GNN）则突破单体单位视角，将事业单位、供应商、审批人等实体建模为节点，通过关系边聚合邻居信息，从而识别通过关联方隐匿的协同舞弊行为^[3]。模型训练过程中，通常采用交叉验证评估泛化能力，并结合贝叶斯优化对超参数进行精细调优，确保模型在未知数据上的稳健表现。

2.5 模型评估与部署

模型评估不能仅依赖整体准确率，尤其在类别极度不平衡的事业单位场景下，准确率可能严重失真。应重

点关注召回率（避免漏报重大风险）、精确率（减少对正常业务的干扰）及其调和平均F1-score，同时辅以AUC-ROC曲线全面衡量判别能力。一旦模型通过严格验证，便可部署至实际应用场景。例如，集成到财政一体化管理平台中，作为预算执行的实时监控模块；或嵌入单位内控系统，对报销申请进行前置风险筛查。部署方式可采用微服务架构或API接口，确保系统响应速度与可扩展性，最终形成“数据驱动、智能预警、人工复核、制度反馈”的闭环 workflow，实现从被动响应向主动防控的转变。

3 主流AI技术在财务舞弊识别中的应用分析

3.1 集成学习模型：XGBoost的卓越表现

在众多人工智能算法中，XGBoost凭借其高效性、鲁棒性与一定的可解释性，成为事业单位财务舞弊识别的标杆模型。其通过引入正则化项控制模型复杂度，有效防止过拟合，同时支持稀疏数据自动处理，能智能学习缺失值的最佳分裂方向。更重要的是，XGBoost内置特征重要性评估机制，可清晰展示哪些变量（如“年末报销激增”“单一供应商占比”）对风险判定贡献最大，便于审计人员追溯关键线索。实证研究表明，在多个高校、医院的试点项目中，XGBoost在F1-score和AUC等指标上显著优于传统方法，尤其在结构化报销与预算数据建模中展现出强大竞争力。

3.2 图神经网络（GNN）：关系视角的突破

传统舞弊识别模型往往将事业单位视为孤立个体，忽略了其嵌入的复杂协作网络。图神经网络（GNN）的引入打破了这一局限。通过将单位、供应商、项目负责人、审批领导等建模为图节点，将采购合同、资金往来、隶属关系等建模为边，GNN能够捕捉风险在关系网络中的传播效应。例如，当某供应商被识别为高风险时，与其频繁交易的多个课题组或科室也可能存在协同虚报行为；又如，同一审批人签批的多笔异常报销，可能指向权力滥用。GNN通过消息传递机制聚合邻居信息，生成的节点嵌入不仅包含自身财务特征，还融合了其在网络中的结构角色，从而显著提升对隐蔽性、系统性舞弊的识别能力。

3.3 大语言模型（LLM）：从文本中洞察风险

财务与业务文档不仅是数据载体，其文字叙述同样蕴含丰富风险信号。大语言模型（LLM）如BERT、ChatGLM等凭借强大的语义理解能力，为挖掘非结构化文本中的舞弊线索提供了新可能。通过对项目申报书、验收报告、报销事由等文本的深度解析，LLM可判断表述是否与财务数据一致，是否存在逻辑矛盾或过度模糊^[4]。例如，某科研项目申报书中承诺购置高端设备，但实际支出集中

于差旅与餐饮,LLM可识别此类“名实不符”;又如,频繁使用“预计”“可能”“相关”等模糊词汇,可能暗示缺乏真实业务支撑。尽管LLM的决策过程仍较难完全解释,但其在文本风险感知上的潜力,使其成为未来多模态舞弊识别系统不可或缺的组成部分。

4 面临的挑战

4.1 数据挑战

尽管AI模型依赖数据驱动,但事业单位在数据层面面临多重瓶颈。一方面,基层单位信息化水平参差不齐,部分凭证仍为纸质,电子化数据缺失或格式混乱;另一方面,财政、税务、银行、工商等系统间存在数据壁垒,难以获取全景视图以追踪资金全链条流向。更严峻的是,舞弊标签存在“假阴性”问题——大量未被发现或未被正式认定的违规行为无法纳入训练集,导致模型学习存在系统性偏差。如何在数据有限、质量不高的现实条件下构建鲁棒、泛化的智能模型,是当前亟待突破的技术与制度难题。

4.2 模型挑战

模型层面的挑战主要集中在可解释性、对抗鲁棒性与动态适应性三个方面。在强调问责与透明的公共部门,深度学习等高性能“黑箱”模型难以满足审计追溯与制度合规要求。其次,随着AI监督普及,舞弊者可能针对性地设计“对抗样本”,如微调报销金额或拆分发票以规避检测规则,这对模型鲁棒性提出更高要求。最后,舞弊手法持续演化,若模型不能及时更新,将面临“概念漂移”风险——即历史学习的模式不再适用于当前环境。因此,开发具备在线学习能力、可解释性强且抗干扰的智能模型,是未来技术演进的关键方向。

4.3 伦理与法律挑战

AI在事业单位财务监督中的广泛应用也引发深刻的伦理与法律思考。算法可能无意中放大对特定类型单位(如科研密集型高校、基层医疗机构)的偏见,导致不公平预警;当AI系统错误地标记一家正常单位为高风险对象时,可能影响其项目申报、资金拨付乃至声誉,此时责任归属尚不明确。此外,在利用人员背景、社交关系等非传统数据时,如何在提升识别精度与保护个人隐私之间取得平衡,亦需严格的法律框架予以规范。这些问题的解决,不仅依赖技术进步,更需制度设计、伦理准则与法律法规的协同演进。

5 结语

人工智能为提升事业单位财务治理能力提供了强大工具,但其应用必须扎根于公共部门的制度逻辑与价值导向。未来的发展方向,不应仅追求技术精度,而应推动“智能模型+内控制度+人员素养+监督问责”的深度融合。唯有如此,才能构建起既高效又可信的财务风险防控体系,切实守护好每一分公共资金,助力事业单位高质量履行公共服务使命。

参考文献

- [1]曾小青,唐湘勇.公司财务舞弊的智能识别与模型优化策略[J].长沙理工大学学报(社会科学版),2021,36(01):81-92.
- [2]梁佳宇,秦诗雨,黄巧琳,等.基于区块链和人工智能技术的企业财务舞弊检测方法研究[J].科技经济市场,2023,(09):28-30.
- [3]陈晓.智能财务时代企业费用报销内控的RPA与舞弊智能识别[J].上海商业,2025,(10):80-82.
- [4]曾凡鹤.业财融合视角下国有企业财务舞弊风险智能防控机制研究[J].国际商务财会,2025,(14):82-85.