

企业财务欺诈异常检测及其最新进展

付秀莲¹, 肖惠予^{2*}

¹中央财经大学管理科学与工程学院, 北京

²北京师范大学政府管理学院, 北京

收稿日期: 2024年2月26日; 录用日期: 2024年3月14日; 发布日期: 2024年4月18日

摘要

财务欺诈对经济市场的稳定性和信任体系造成了重大威胁, 准确识别和检测财务欺诈异常变得至关重要。本文综述了企业财务欺诈异常检测的最新进展。首先介绍了传统的财务欺诈检测方法, 重点讨论了基于机器学习和数据挖掘技术的财务欺诈检测方法。这些方法利用丰富的财务数据和模式识别算法, 可以自动发现潜在的欺诈行为。接着介绍了出现的新方法: 将财务报表、媒体新闻、企业公告等文本数据纳入检测范围。新方法通过提取更多的特征信息、融合多种数据源的方式, 提高了财务欺诈检测的准确性和效率。

关键词

财务欺诈, 异常检测, 机器学习, 数据挖掘, 文本数据

Corporate Financial Fraud Anomaly Detection and Its Latest Progress

Xiulian Fu¹, Huiyu Xiao^{2*}

¹School of Management Science and Engineering, Central University of Finance and Economics, Beijing

²School of Government, Beijing Normal University, Beijing

Received: Feb. 26th, 2024; accepted: Mar. 14th, 2024; published: Apr. 18th, 2024

Abstract

Financial fraud poses a major threat to the stability and trust system of the economic market, so it is very important to accurately identify and detect financial fraud anomalies. This paper reviews the latest progress in anomaly detection of financial fraud in enterprises. Firstly, the traditional financial fraud detection methods are introduced, and the financial fraud detection methods based on

*通讯作者。

文章引用: 付秀莲, 肖惠予. 企业财务欺诈异常检测及其最新进展[J]. 现代管理, 2024, 14(4): 566-573.

DOI: 10.12677/mm.2024.144069

machine learning and data mining technology are mainly discussed. These methods use rich financial data and pattern recognition algorithms to automatically spot potential fraud. Then, a new method is introduced: text data such as financial statements, media news and corporate announcements are included in the scope of detection. The new method improves the accuracy and efficiency of financial fraud detection by extracting more characteristic information and integrating multiple data sources.

Keywords

Financial Fraud, Anomaly Detection, Machine Learning, Data Mining, Text Data

Copyright © 2024 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

狭义上, 企业财务欺诈是指企业为了欺骗投资者、债权人等相关方故意伪造、篡改或隐瞒财务报表信息的行为; 广义上, 企业财务欺诈包括任何企业在财务报告中故意提供虚假或误导性信息的行为, 不仅限于财务报表的篡改, 还包括非法转移资金、操纵市场价格等欺诈行为。在投资者的角度上, 财务欺诈是指企业在财务报表中故意提供虚假信息导致投资决策的错误或损失。尽管这些定义在细节上有所不同, 但它们都强调了企业故意提供虚假或误导性信息的行为, 以及对利益相关方的损害。无论采用何种定义, 准确识别和检测企业财务欺诈都是维护金融市场稳定和投资者利益的重要任务。

为了保护投资者利益、维护市场秩序、防范企业风险, 财务欺诈检测问题一直为学术界、企业界以及政府管理部门所重视。目前企业财务欺诈检测的方法路径主要有两个: 基于财务指标的财务欺诈检测和基于文本信息的财务欺诈检测。

本文对企业欺诈领域的文献进行了综合梳理, 系统地总结了当前研究的主要成果和发展趋势。通过对不同研究方法和模型的综合比较, 本文揭示了企业欺诈检测领域的研究现状, 并提出了未来研究的重要方向和挑战。本综述的边际贡献在于对先前综述文献所未涵盖的样本范围进行扩展, 为研究者提供全面的研究视角, 促进学术交流和跨学科合作, 为进一步深化企业欺诈检测研究提供有益的启示。

2. 基于财务数据指标的财务欺诈检测

财务报表中包含公司的资产、负债、权益、收入、支出、现金流等财务信息, 是公司用来检测财务业绩并做出合理决策的有效工具, 同时也是社会、政府监管部门和投资者衡量企业经营状况的重要工具。基于财务指标进行的财务欺诈检测是依据公司出具的财务报表数据进行的。

2.1. 财务指标的选取

财务报表中的数据指标是企业财务和运营状况的综合体现, 企业财务欺诈行为在财务报表中必然有所体现, 具体表现形式为财务结构和财务指标的异常。指标的历史变化以及指标间的相关性对于欺诈检测有着重要的影响, 因此在进行企业财务欺诈检测时要进行全面、合理、具体的财务指标选取。黄学敏首次从行业间横向和企业自身纵向比较两个维度出发, 通过 LPM 模型和 Logit 模型回归分析发现, 6 个基于期间费用、折旧、财务损益和应收账款的财务指标统计显著性较高, 可以作为预警指标[1]。杨芳等人首次从实证角度提出, 在财务欺诈识别模型中引入现金流量指标, 可以提高模型的识别能力[2]。袁先

智等人则以吉布斯随机搜索算法对上市公司的财务报表数据进行分析, 解决了由于财务报表勾稽产生的高维数据灾难问题, 并提取出八个可以刻画上市公司财务欺诈风险的关联特征因子[3]。陈孝新从财务杠杆、权责发生制、资产盈利能力三方面进行考量, 选择出资产负债率、资产毛利率、资产周转率和运营资金比率四个模型显著性比率, 将检验准确率提升至 93.33% [4]。

以上研究注重检测指标的选择, 忽略了企业在时间序列上的表现。余玉苗关注到动态指标对于鉴别财务舞弊行为的有利性, 从欺诈行为发生前一年与当年财务指标的动态增量视角入手, 发现固定资产增长率、经营现金流量对流动负债比率等五个财务指标的变动对财务舞弊产生重要影响[5]。熊芮卿提炼使用时序指标, 提升了模型的识别精度, 并且证明比值指标优于差值和相对值指标, 较近年份的时序指标影响力更大[6]。

企业欺诈检测指标选择经历了由静态向动态拓展的演变过程。动态指标的引入提升了模型的识别准确性, 有助于更全面地评估企业的财务状况和发展态势。

2.2. 模型的选择

常见欺诈的类型分为四类: 银行欺诈、财务报表欺诈、保险欺诈和加密货币欺诈。据 Al-Hashedi 等人调查, 共有 34 种数据挖掘技术被用于识别各种金融欺诈行为, 大多数技术被用于银行欺诈和保险欺诈[7]。其中, 支持向量机(SVM)、神经网络(NN)和随机森林(RF)是使用最为广泛的金融欺诈检测技术。从 1995 年开始, Persons 首先利用 Logistic 模型从初选的各项财务指标中选取了与欺诈相关的 4 项指标[8], 之后便不断有人试图将各种大数据挖掘技术应用于财务报表欺诈检测领域, 各种大数据挖掘技术得到了创新与发展。

2.2.1. 支持向量机(Support Vector Machine)

支持向量机(Support Vector Machine, SVM)是一种基于统计学习理论的监督学习算法, 广泛应用于模式识别和机器学习领域。其主要目标是在高维空间中找到一个最优的超平面, 能够对不同类别的样本进行最佳的分类。

SVM 具有较强的泛化能力, 可有效处理高维数据。邓青山首次采用 SVM 构建财务欺诈检测模型, 利用 6 个财务指标进行训练和验证, SVM 的平均准确率达到 87.79%, 成功验证了其在财务欺诈检测方面的有效性[9]。

通过引入核函数, SVM 能够有效地处理非线性分类问题。李秀智和应双双采用 Lib-SVM 算法, 结合 RBF 核和线性核构建了一种用于检测调节利润财务报表欺诈的模型, 在预测准确率方面达到了 86.67%, 明显高于 Logistic 回归模型的表现水平。这表明该模型在识别财务报表欺诈方面具有较高的潜力和准确性[10]。

此外, SVM 在预测性能方面表现良好, 并且在小样本数据集具有较好的性能。Moepya 等人对比了 SVM 模型与 k 最近邻(kNN)方法和 Logistic 回归(LR), 结果显示, SVM 方法提供了更高的分类准确性[11]。Yao 等人将 SVM、KNN、分类与回归树(CART)、反向传播神经网络(BP-NN)、逻辑回归(LR)和贝叶斯分类器(Bayes)进行对比, 发现 SVM 在所有条件下都具有最高的准确率[12]。El-Bannany 等人将 SVM、LR、NN 和决策树(DT)应用于阿拉伯联合酋长国(UAE)公司财务报表欺诈检测, SVM 的准确率为 89.54%, 优于其他分类器[13]。Chi 等人将 SVM、决策树 C5.0 和 CHAID 模型相互结合应用于财务报告欺诈的检测, 发现 SVM 与其他模型相结合形成的模型具有较好的预测性能[14]。

SVM 在检测企业财务欺诈问题上也存在一些挑战和限制。对于大规模数据集, 支持向量机的计算复杂度较高, 需要较长的训练时间和更多的计算资源。SVM 核函数的选择需要一定的经验和领域知识, 选择不当可能影响模型性能。支持向量机对于噪声和异常点比较敏感, 可能导致模型过度拟合或者性能下降。

2.2.2. 神经网络(Neural Network)

神经网络(Neural Network, NN)又称人工神经网络(Artificial Neural Network, ANN), 是一种基于生物神经元结构和功能而发展起来的计算模型, 通过采用反向传播算法(Back Propagation, BP)比较网络的输出和目标输出之间的误差并最小化误差, 来解决分类、回归、聚类、图像和语音识别等各种问题。神经网络具有可处理非线性问题、自适应性强、并行处理能力强和可泛化能力强的优点, 在欺诈检测应用性较强。

Fanning 等人于 1998 年首次利用 ANN 深入检查财务报表, 并发现 ANN 是一个具有高概率检测虚假财务报表的模型[15]。Rizki 等人应用 SVM 和 ANN 进行财务欺诈检测, 发现 ANN 对没有特征选择的数据产生最高的准确率, 为 90.97% [16]。Lin 等人将人工神经网络与逻辑回归和决策树(CART)对比应用于分类测试, 发现 ANN 对训练样本和测试样本的正确分类率分别为 91.2%和 90.4%, 均高于 CART 方法和 Logistic 模型[17]。Nawaiseh 等人探讨了基于支持向量机(SVM)、人工神经网络(ANN)和 K 最近邻(KNN)作为财务报表审计预测分类模型的数据挖掘能力, 实证结果表明, 人工神经网络和支持向量机技术在非欺诈公司正确分类方面取得了更高的平均准确率[18]。范斌等人针对财务凭证多类型数据(文本数据、类别型数据和数值型数据)和数据正负样本比例失衡问题, 基于代价敏感学习对多分支深度神经网络算法进行了改进, 有效地缩小了审查范围, 而且模型在负样本数据不断丰富的情況下准确度也实现不断提升[19]。人工神经网络与逻辑回归、支持向量机、K 近邻等机器学习算法相比综合性能较好, 在财务欺诈检测中具有较高的检测精度。

2.2.3. 集成算法(Ensemble Method)

集成算法是一种机器学习技术, 旨在通过将多个基本学习器(通常是弱学习器)的预测结果进行组合, 以获得更准确、鲁棒的预测或分类结果。与传统机器学习算法相比, 集成算法可以减少过拟合和欠拟合现象的发生, 可以修正模型的错误, 具有较强的鲁棒性和预测性能。常见的集成算法有袋装法(Bagging)即随机森林、提升法(Boosting)包括 AdaBoost 和 Gradient Boosting 和堆叠法(Stacking)。其中, 常见的 Gradient Boosting 算法有 XGBoost 和 LightGBM。

Perols 比较了六种流行的统计和机器学习模型的表现, 使用堆叠分类器检测财务欺诈, 并与单个分类比较[20]。然而, 该研究在组合模型时没有考虑基分类器的多样性, 导致得到令人惊讶的 bagging 和堆叠分类器次优结果。Song 等人通过四个分类器(逻辑回归、反向传播神经网络、C5.0 决策树和支持向量机)和这些分类器的集合进行实验, 实验结果表明, 集合分类器在准确性和复合错误率方面分别优于单个分类器[21]。此外, 在真阳性率即欺诈公司被正确归类为欺诈公司方面, 集成方法优于其他方法[22]。

企业财务欺诈检测问题具有高度不平衡性, Ye 等人使用 SMOTE 进行采样检测, 采用随机森林来学习不平衡数据, 在 11726 份财务报表中, 随机森林的检测准确度优于 ANN、LR、SVM、CART 等其他算法[23]。Liu 等人将集成模型与 SVM、LR 等传统模型进行对比, 通过比较两者的 AUC 和 KS 值, 发现集成模型 XGBoost、LightGBM 和 RF 的识别能力更强, 此外还使用 AUC 值最高的 Smote 过采样算法处理样本数据集不平衡问题[24]。

Ali 等人使用 SMOTE 解决样本数据类不平衡问题, 利用集成算法 XGBoost 构建财务欺诈检测模型, 并和 LR、DT、SVM、AdaBoost 以及 RF 五种机器学习技术一同进行预测财务欺诈, 实证结果表明, 优化之后的 XGBoost 的预测结果最优, 模型检测准确率最高[25]。同样, Khedr 等人也在识别来自中东和北非地区的一组样本公司的欺诈行为过程中得出了相同的结论[26]。Achakzai 等人使用动态集成选择技术(DES)构建了一个集成模型, 与静态集成学习相比, DES 在运行时会根据样本数据和其他因素的变化动态地组合一组基本分类器的预测结果作为最终预测, 以提高模型的泛化性能和鲁棒性[27]。

3. 基于文本信息的财务欺诈检测

对上市公司欺诈识别的研究, 经历了从简单到复杂的过程, 研究路径由简单的模型分析到建立回归分析再到复杂模型的应用最终到采用混合分类器的集成算法进行研究。这些研究局限于对结构性财务数据进行识别, 审查公司在已经过去的季度或者年份的表现, 研究具有一定的时间滞后性。利用社交媒体平台实时产生的丰富内容可以极大地解决财务报表的滞后性问题, 这需要运用到文本挖掘技术。

文本挖掘技术是一个集普通语言处理、信息检索、web 挖掘、计算语言学、数据提取和数据挖掘为一体的跨学科领域, 具有很高的商业价值。数据性指标由于结构化程度高、精确度和准确性高、可量化等优点被广泛应用于企业财务欺诈识别研究领域。而文本数据却能提供更加综合和全面的信息, 与各种指标结合起来, 可以更好第体现数据所反映的现实问题。文本数据作为一种非结构化数据, 在金融和会计分析领域的应用已经被证实具有重要作用, 主要集中应用于信息品质和数量的分析、信息披露和市场异常的研究, 以及构建企业性能指标的分析。这些应用场景包括但不限于交易监测和异常检测、舆情监测、内部调查和举报管理、合规监督和报告, 还有舆情预警和风险管理。这些应用是文本数据分析的重要工具, 并且具有广泛的适用范围。

3.1. 文本数据来源

文本挖掘技术可以对不同的数据源进行企业欺诈识别。公司的财务报表是最直接的来源之一, 其中包含了公司的财务状况、经营活动和盈利能力等信息。这些报表中的文字描述、注释、管理层讨论和分析部分都可能含有关键信息, 可以用于进行文本挖掘和分析。

公司相关的新闻报道、行业分析、舆情监控等媒体信息都可以作为文本数据来源。这些信息可能包括对公司经营活动、产品、服务、管理层及员工的评价和看法, 以及市场反应等内容。新闻报道可以真实地反映出大众对公司的评价, 这些评价体现出的情感倾向可以较为真实地反映出上市公司以及非上市公司的经营状况和盈利水平。以新闻报道等媒体信息作为数据源打破了只能对上市公司进行财务欺诈检测的限制。此外, 通过社交媒体平台、在线论坛、博客评论等渠道, 收集关于公司或其产品、服务的用户评论、投诉、赞扬等信息也可以作为文本数据源。

另一个比较重要的数据源便是企业公告。上市公司主要通过公告来披露文本信息, 主动披露有关事项可以减轻公司管理层和信息需求者之间的信息不对称问题。为了让外界全面了解公司情况, 通常公告会包含多方面内容, 如业绩报告反映公司运营状况、重大事项和风险提示, 以及停牌复牌的交易提示等。

3.2. 文本挖掘在财务欺诈检测的应用

虚假财务报表和非虚假财务报表在主/被动语态结构、语气词等语言特征方面存在差异。有研究表明, 年度财务报告中的财务指标信息隐藏的欺诈细节可以由报告中的语言线索进行补充, 以此评估年报的真实性和诚实程度。财务异常识别的文本挖掘技术集中应用在财报的 MD&A 部分, 文本挖掘的应用缺乏一个系统的、整体的、理论的分析框架来指导欺诈检测工作。董伟将主题、观点、情感、情态、人称代词、写作风格和体裁七种信息类型提取词级特征和文档级特征用于线性支持向量机分类器的输入, 在十倍交叉验证下, 该分析框架的分类平均准确率为 82.36%, 优于使用财务比率的基线方法[28]。公司财务报表的定性信息提供了有用的信息, 可以提高欺诈预测模型的准确性[29]。程双双等人以财务报表中的 MD&A 为文本数据, 采用长文本分类模型 XLNet_HAN 进行财务欺诈识别检测, 该模型在处理非平衡特性的文本数据方面具有更好的性能[30]。

针对财务报表数据识别的时滞性, 李华玲等人通过爬取东方财富网和金融界股票频道新闻库中的实时新闻进行了模型识别, 提取了文本数据中的有效隐含信息, 提高了识别企业欺诈行为的及时性[31]。现

阶段单纯的文本数据研究较少, 研究人员更倾向于文本信息和财务指标信息相结合, 以提高检测准确度。

为了比较文本信息和财务指标信息对财务造假识别的影响, 有人对比了基于财务指标、基于新闻情感、基于财务指标和新闻情感的三种上市公司财务造假 Logistic 识别模型的性能差异, 发现综合新闻情感与财务指标的模型准确率最高。Craja 等人利用分层注意力网络(HAN)从 MD&A 中提取文本特征与财务比率信息相结合, 发现文本特征提供了更有希望的分类结果, 并大大增强了财务比率的检验效果[32]。刘会醒等人利用 Jieba 分词从文本语调、相似度和复杂度三个角度构建了文本指标, 将 MD&A 中文文本信息和财务指标信息加入到 Stacking 集成算法模型中, 分类器的识别性能在准确率和召回率方面得到了显著的提升[33]。

此外, 考虑到指标在时间轴上的变化, 陈朝焰等人在财务指标数据中融入财经新闻文本特征并提取时序信息, 将准确率提高了 8%, 为数据特征提供了有效的补充[34]。李爱华和肖庆兰则通过融合多种数据源的方式再次验证了文本特征对于财务欺诈检测的有效性[35] [36]。

文本数据的引入拓宽了研究对象和研究范围, 弥补了财务指标数据在时间上的滞后性, 增强了数据的真实性, 但同时也存在一些问题。第一, 文本信息本身不够明确, 需要研究者进行加工处理才能用于计量分析, 这可能引入噪音和错误, 并且数据处理的可重复性存在争议。第二, 大部分文本数据缺乏权威来源, 数据的可靠性有待考量。第三, 由于文本数据量庞大, 现有的计量分析方法并不一定适用。第四, 应用文本大数据分析技术研究会计和金融问题对研究者的综合能力, 尤其是编程和数量分析能力提出了挑战。这些问题都有待进一步解决。

4. 最新研究进展与结论

目前, 利用财务报表进行欺诈检测的研究已经涵盖了各种机器学习方法, 对于文本信息的研究正逐步从以 MD&A 为中心向其他信息源拓展, 对于信息的横向和纵向特征的选取上也有了更多的考量。与此同时, 对于财务欺诈检测新模型、新方法和新指标的研究并未停歇。在模型的创新上, Islam 等人针对数据集高度不平衡问题, 开发了一个不涉及任何重采样技术的欺诈交易检测规则——基于规则的模型 (RBM), 模型性能优于现有的机器学习模型[37]; Innan 等人提出了使用量子图神经网络(QGNN)检测金融欺诈的新方法, 使用变分量子电路(VQC)提高 QGNN 的性能, 使其可以处理图结构数据[38]; Wu 等人基于异构图神经网络提出了一个多任务学习框架——MultiFraud, 用于复杂的欺诈检测和提供合理解释, 实现欺诈检测跨领域执行和解释[39]。为了确定审计师之间财务欺诈的风险概况, Munteanu 等人应用多对应因子分析(AFCM), 建议将财务指标分为四类: 盈利能力指标、融资和负债指标、流动性指标和业务活动报告[40]。财务欺诈问题一直为社会和群众所重视, Kanaparthi 发现从 2017 年到 2022 年, 财务欺诈领域的出版物显著增加, 这凸显了人们对该领域日益增长的兴趣和不断扩大的研究活动, 也表明了该领域在当代学术领域的潜在意义和相关性[41]。

本文基于企业财务欺诈文献梳理发现, 对于企业财务欺诈识别的研究主要集中于基于财务报表指标检测、优化模型和解决不平衡样本问题, 将公司类型限制在上市公司范围内, 忽略了财务报表在时间上的滞后性。文本挖掘技术引入的文本信息补充了更多的细节信息, 有效解决了基于财务指标欺诈检测的滞后性, 同时也将非上市公司纳入到可进行欺诈检测领域, 拓宽了研究对象和范围。文本信息的应用在信息的处理、数据质量的提升、与现有机器学习模型的契合度以及与其他信息源的融合等方面还没有得到有效的进步, 有待进一步研究。

参考文献

- [1] 黄学敏. 上市公司财务欺诈及其监管研究[D]: [博士学位论文]. 厦门: 厦门大学, 2006.

- [2] 杨芳, 刘端, 汪子文. 现金流量指标在上市公司财务欺诈识别中应用的实证研究[J]. 金融经济(理论版), 2006(8): 122-124.
- [3] 袁先智, 周云鹏, 严诚幸, 等. 公司财务欺诈预警与风险特征筛选的新方法: 基于人工智能算法[C]//中国管理现代化研究会, 复旦管理学奖励基金会. 第十五届(2020)中国管理学年会论文集: 2020 年卷. 北京: 中国管理现代化研究会, 2020: 1-16.
- [4] 陈孝新. 上市公司财务欺诈的识别模型[J]. 统计与决策, 2005(13): 44-45.
- [5] 余玉苗, 吕凡. 财务舞弊风险的识别——基于财务指标增量信息的研究视角[J]. 经济评论, 2010(4): 124-130.
- [6] 熊芮卿. 基于时序性信息的财务报表欺诈识别[D]: [硕士学位论文]. 成都: 西南交通大学, 2013.
- [7] Al-Hashedi, K.G. and Magalingam, P. (2021) Financial Fraud Detection Applying Data Mining Techniques: A Comprehensive Review from 2009 to 2019. *Computer Science Review*, **40**, Article ID: 100402. <https://doi.org/10.1016/j.cosrev.2021.100402>
- [8] Persons, O.S. (1995) Using Financial Statement Data to Identify Factors Associated with Fraudulent Financial Reporting. *Journal of Applied Business Research (JABR)*, **11**, 38-46. <https://doi.org/10.19030/jabr.v11i3.5858>
- [9] Deng, Q. (2009) Application of Support Vector Machine in the Detection of Fraudulent Financial Statements. 2009 4th International Conference on Computer Science & Education, Nanning, 25-28 July 2009, 1056-1059. <https://doi.org/10.1109/ICCSE.2009.5228542>
- [10] Li, X. and Ying, S. (2010) Lib-SVMs Detection Model of Regulating-Profits Financial Statement Fraud Using Data of Chinese Listed Companies. 2010 International Conference on E-Product E-Service and E-Entertainment, Henan, 7-9 November 2010, 1-4. <https://doi.org/10.1109/ICEEE.2010.5660371>
- [11] Moepya, S.O., Nelwamondo, F.V. and Van Der Walt, C. (2014) A Support Vector Machine Approach to Detect Financial Statement Fraud in South Africa: A First Look. *Intelligent Information and Database Systems: 6th Asian Conference, ACIIDS 2014*, Bangkok, 7-9 April 2014, 42-51. https://doi.org/10.1007/978-3-319-05458-2_5
- [12] Yao, J., Pan, Y., Yang, S., et al. (2019) Detecting Fraudulent Financial Statements for the Sustainable Development of the Socio-Economy in China: A Multi-Analytic Approach. *Sustainability*, **11**, Article 1579. <https://doi.org/10.3390/su11061579>
- [13] El-Bannany, M., Dehghan, A.H. and Khedr, A.M. (2021) Prediction of Financial Statement Fraud Using Machine Learning Techniques in UAE. 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), Monastir, 22-25 March 2021, 649-654. <https://doi.org/10.1109/SSD52085.2021.9429297>
- [14] Chi, D.J., Chu, C.C. and Chen, D. (2019) Applying Support Vector Machine, C5.0, and CHAID to the Detection of Financial Statements Frauds. *Intelligent Computing Methodologies: 15th International Conference, ICIC 2019*, Nanchang, 3-6 August 2019, 327-336. https://doi.org/10.1007/978-3-030-26766-7_30
- [15] Fanning, K.M. and Cogger, K.O. (1998) Neural Network Detection of Management Fraud Using Published Financial Data. *Intelligent Systems in Accounting, Finance & Management*, **7**, 21-41. [https://doi.org/10.1002/\(SICI\)1099-1174\(199803\)7:1<21::AID-ISAF138>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1099-1174(199803)7:1<21::AID-ISAF138>3.0.CO;2-K)
- [16] Rizki, A.A., Surjandari, I. and Wayasti, R.A. (2017) Data Mining Application to Detect Financial Fraud in Indonesia's Public Companies. 2017 3rd International Conference on Science in Information Technology (ICSITech), Bandung, 25-26 October 2017, 206-211. <https://doi.org/10.1109/ICSITech.2017.8257111>
- [17] Lin, C.C., Chiu, A.A., Huang, S.Y., et al. (2015) Detecting the Financial Statement Fraud: The Analysis of the Differences between Data Mining Techniques and Experts' Judgments. *Knowledge-Based Systems*, **89**, 459-470. <https://doi.org/10.1016/j.knosys.2015.08.011>
- [18] Nawaiseh, A.K., Abbod, M.F. and Itagaki, T. (2020) Financial Statement Audit Using Support Vector Machines, Artificial Neural Networks and K-Nearest Neighbor: An Empirical Study of UK and Ireland. *International Journal of Simulation—Systems, Science & Technology*, **21**, 1-6. <https://doi.org/10.5013/IJSSST.a.21.02.07>
- [19] 范斌, 宁德军, 卢俊哲, 等. 基于加权 KNN 与代价敏感多分支深度神经网络的审计数据异常检测[J]. 计算机应用与软件, 2024, 41(2): 100-108.
- [20] Perols, J. (2011) Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. *Auditing: A Journal of Practice & Theory*, **30**, 19-50. <https://doi.org/10.2308/ajpt-50009>
- [21] Song, X.P., Hu, Z.H., Du, J.G., et al. (2014) Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China. *Journal of Forecasting*, **33**, 611-626. <https://doi.org/10.1002/for.2294>
- [22] Hajek, P. and Henriques, R. (2017) Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud—A Comparative Study of Machine Learning Methods. *Knowledge-Based Systems*, **128**, 139-152. <https://doi.org/10.1016/j.knosys.2017.05.001>
- [23] Ye, H., Xiang, L. and Gan, Y. (2019) Detecting Financial Statement Fraud Using Random Forest with SMOTE. *IOP Con-*

- ference Series: Materials Science and Engineering*, **612**, Article ID: 052051. <https://doi.org/10.1088/1757-899X/612/5/052051>
- [24] Liu, Z., Ye, R. and Ye, R. (2021) Detecting Financial Statement Fraud with Interpretable Machine Learning. <https://doi.org/10.21203/rs.3.rs-640038/v1>
- [25] Ali, A.A., Khedr, A.M., El-Bannany, M., *et al.* (2023) A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. *Applied Sciences*, **13**, Article 2272. <https://doi.org/10.3390/app13042272>
- [26] Khedr, A.M., El Bannany, M. and Kanakkayil, S. (2021) An Ensemble Model for Financial Statement Fraud Detection. *Journal of Universal Computer Science*, **2**, e69590. <https://doi.org/10.3897/arphapreprints.e69590>
- [27] Achakzai, M.A.K. and Peng, J. (2023) Detecting Financial Statement Fraud Using Dynamic Ensemble Machine Learning. *International Review of Financial Analysis*, **89**, Article ID: 102827. <https://doi.org/10.1016/j.irfa.2023.102827>
- [28] Dong, W., Liao, S. and Liang, L. (2016) Financial Statement Fraud Detection Using Text Mining: A Systemic Functional Linguistics Theory Perspective. *Proceeding of the 20th Pacific Asia Conference on Information Systems*, Chiayi, June 27-July 1 2016, 188.
- [29] Nießner, T., Gross, D.H. and Schumann, M. (2022) Evidential Strategies in Financial Statement Analysis: A Corpus Linguistic Text Mining Approach to Bankruptcy Prediction. *Journal of Risk and Financial Management*, **15**, Article 459. <https://doi.org/10.3390/jrfm15100459>
- [30] 程双双, 谷晓燕, 王兴芬. 基于非平衡MDA文本数据的财务欺诈识别[J]. 管理现代化, 2024(1): 121-127.
- [31] 刘华玲, 陶龙, 曹亚珂, 等. 基于文本数据挖掘的上市公司财务欺诈识别研究[C]//中国管理现代化研究会, 复旦管理学奖励基金会. 第十八届(2023)中国管理学年会暨“一带一路”十周年研讨会论文集: 2023年卷. 上海: 上海对外经贸大学统计与信息学院, 2023: 1-8.
- [32] Craja, P., Kim, A. and Lessmann, S. (2020) Deep Learning for Detecting Financial Statement Fraud. *Decision Support Systems*, **139**, Article ID: 113421. <https://doi.org/10.1016/j.dss.2020.113421>
- [33] 刘会醒, 程建华. 基于集成学习和文本分析的财务欺诈识别研究[J]. 福建商学院学报, 2023(4): 42-52.
- [34] 陈朝焰, 韩冬梅, 吴馨一. 融合新闻文本和时序信息的上市公司财务欺诈预警[J]. 财会月刊, 2023, 44(12): 30-39.
- [35] 李爱华, 王迪文, 续维佳, 等. 基于多数据源融合的创业板上市公司财务造假异常检测[J]. 数据分析与知识发现, 2023, 7(5): 33-47.
- [36] 肖庆兰. 财务指标与文本相结合的上市公司财务欺诈识别[D]: [硕士学位论文]. 南昌: 南昌江西财经大学, 2022.
- [37] Islam, S., Haque, M.M. and Karim, A.N.M.R. (2024) A Rule-Based Machine Learning Model for Financial Fraud Detection. *International Journal of Electrical & Computer Engineering*, **14**, 759-771. <https://doi.org/10.11591/ijece.v14i1.pp759-771>
- [38] Innan, N., Sawaika, A., Dhor, A., *et al.* (2024) Financial Fraud Detection Using Quantum Graph Neural Networks. *Quantum Machine Intelligence*, **6**, Article No. 7. <https://doi.org/10.1007/s42484-024-00143-6>
- [39] Wu, B., Chao, K.M. and Li, Y. (2024) Heterogeneous Graph Neural Networks for Fraud Detection and Explanation in Supply Chain Finance. *Information Systems*, **121**, Article ID: 102335. <https://doi.org/10.1016/j.is.2023.102335>
- [40] Munteanu, V., Zuca, M.R., Horaicu, A., *et al.* (2024) Auditing the Risk of Financial Fraud Using the Red Flags Technique. *Applied Sciences*, **14**, Article 757. <https://doi.org/10.3390/app14020757>
- [41] Kanaparthi, V. (2024) Transformational Application of Artificial Intelligence and Machine Learning in Financial Technologies and Financial Services: A Bibliometric Review. *International Journal of Engineering and Advanced Technology*, **13**, 71-77. <https://doi.org/10.35940/ijeat.D4393.13030224>