

# 企业技术风险阈值激活模型构建研究

## ——以智能网联汽车产业链为例

张 越 曹 悅 白 晨

(中国科学技术信息研究所, 北京 100038)

**摘要:** 响应强化产业风险监测能力的实际管理需求, 提升管理部门和智库研究人员对特定产业中企业技术风险问题的分析能力, 为技术风险激活或抑制的规律性认识提供机器学习模型支撑。首先, 对企业风险的定义和多维度类型进行划分, 采用基于机器学习的方法构建企业技术风险阈值激活模型, 深入挖掘企业风险特征。其次, 运用随机森林、XGBoost等8种机器学习算法训练参数变量, 以学习企业技术风险的属性, 并对模型有效性进行评价。再次, 以智能网联汽车产业链为例利用自动化方法揭示企业风险激活的规律性特征, 3种梯度提升综合模型的分类预测准确率达到82.59%, 实现从大量相关数据变量中识别出具有潜在技术风险的企业。最后, 提出未来的研究将是进一步提高模型的预测精度及稳定性, 并加强其在其他领域企业技术风险评估中的应用。

**关键词:** 企业技术风险; 阈值激活; 分类预测; 数据挖掘; 智能网联汽车

**DOI:** 10.3772/j.issn.1674-1544.2024.03.001

**CSTR:** 15994.14.issn.1674.1544.2024.03.001

中图分类号: F204

文献标识码: A

## Research on Enterprise Technical Risk Threshold Activation Model Construction

### ——Take Intelligent Connected Vehicle (ICV) Industrial Chain as an Example

ZHANG Yue, CAO Yue, BAI Chen

(Institute of Scientific and Technical Information of China, Beijing 100038)

**Abstract:** In response to the practical management needs for strengthening industrial risk monitoring capabilities, this approach enhances the analytical abilities of management departments and think tank researchers in analyzing enterprise technology risk issues within specific industries. It also provides support through machine learning models for a systematic understanding of technology risk activation or suppression. By defining enterprise risk and categorizing it into multiple dimensions, a machine learning-based method is utilized to construct an Enterprise Technology Risk Threshold Activation Mode, which aims to deeply explore the characteristics of enterprise risk. It employs eight machine learning algorithms, including Random Forest, XGBoost, etc., to train parameter variables for learning the attributes of enterprise technology risk, as well as for evaluating the model's effectiveness. It reveals the regularity characteristics of corporate risk activation using automated methods in the case of the Intelligent Connected Vehicle (ICV) industrial chain, with the classification prediction accuracy of three gradient boosting synthesis models reaching 82.59%. This enables the

**作者简介:** 张越 (1994—), 男, 中国科学技术信息研究所助理研究员, 研究方向为情报学 (通信作者); 曹悦 (1994—), 女, 中国科学技术信息研究所助理研究员, 研究方向为科技资源管理; 白晨 (1980—), 女, 中国科学技术信息研究所副研究员, 研究方向为科技资源管理。

**基金项目:** 中国科学技术信息研究所创新研究基金青年项目“产业风险阈值激活模型构建研究” (QN2023-03)。

**收稿时间:** 2023年11月14日。

identification of enterprises with potential technical risks from a large dataset of relevant variables. Future work will focus on further improving the prediction accuracy and stability of these models, as well as expanding their application in enterprise technology risk assessment across various fields.

**Keywords:** enterprise technology risk, threshold activation, classification prediction, data mining, intelligent connected vehicle(ICV)

## 0 引言

当今世界，国际局势的不确定性使得我国产业风险日益严峻。针对部分战略性新兴产业，技术研发不确定性、新兴技术治理挑战、资源断供、技术歧视与垄断等多维风险不断涌现。为了响应党中央对科技安全工作的要求，以及加强产业风险的监测能力的实际管理需求，提升管理部门、智库研究人员对产业聚焦断点、薄弱环节、“卡脖子”风险问题的分析能力，利用工具化模型化能力识别产业所面临的风险因素，加强产业风险预警、确保产业安全尤为紧迫。风险阈值激活作为用于管理风险中感知风险环节的一部分，其实实践对于成功的战略预警、风险管理等预见性活动至关重要。本文的研究结果将为特定产业中企业技术风险激活或抑制的规律性认识提供机器学习模型支撑，丰富研究视角，帮助产业前瞻性地识别潜在风险、增强风险抵御能力，为产业监管政策制定、新型产业体系中资源合理配置提供一定借鉴。

## 1 研究现状

### 1.1 文献综述

已掌握的文献显示，国内外学者对产业技术创新风险开展的相关研究主要归纳为以下3类。

一是对具体风险介绍及风险应对措施运用的阐释性分析。这类研究侧重于详细识别和梳理各种与企业技术创新相关的风险类型，如市场风险、技术实施风险、知识产权侵犯风险以及合作项目风险等。宋珊珊等<sup>[1]</sup>分析认为，管理技术创新风险是确保企业技术创新进程成功的重要环节。这一过程包括环境信息的搜集、风险的识别、对风险的评估以及风险决策等关键步骤。有

效地管理风险可以减少不确定性，进而推动技术创新的顺利实施。郝玲玲<sup>[2]</sup>深入分析了摩托罗拉公司的铱星系统项目，探讨其创新过程中的挑战与技术创新失败的关键因素，从技术哲学视角重新审视了技术创新伴随的风险性问题。朱爱孔等<sup>[3]</sup>分析棉纺织产业的内外部环境，对影响其可持续发展的限制因素进行了识别，并基于危机性产业衰退理论，提出了优化棉花产业布局、改进危机风险应对措施以及建立产业风险预警机制的策略。

二是从风险管理的框架原理出发，有学者采用一定的指标体系、风控模型、情景分析等方法来解析风险传播、扩散的概念化和程式化现象。雷志梅<sup>[4]</sup>探讨了产业经济复杂网络结构中的经济风险扩散模型，有效揭示了产业经济风险扩散的动态属性以及其传播途径。杨君良<sup>[5]</sup>构建事前感知策略与事后自适应策略的产业共生网络风险传播模型，分析在邻居发生风险前后组织的行为对共生网络中风险传播的作用。在具体行业风险预警模型的可行性验证与应用方面，林金枫<sup>[6]</sup>采用风险情景识别方法，从风险源角度分析了中国新能源汽车产业的风险，并利用FIR（因素—交互行为—风险）模型探究了各因素间的交互关系，以此识别风险，进而构建了新能源汽车产业风险的多维图景。王粟旸等<sup>[7]</sup>通过基于极值理论的VAR模型和CD模型，对银行业和房地产业各机构的风险水平进行了评估，并研究了这些行业内外的风险传播情况。这些研究为不同行业的风险管理提供了深刻的方法支持。

三是聚焦于风险阈值研究领域，利用相关算法模型等手段，探讨风险管理领域中风险阈值的定义及其确定方法。国内风险阈值研究的提出是在2010年前后。夏皓<sup>[8]</sup>指出风险阈值是风险源、

风险载体、风险流、风险表征、传导路径之外影响风险传导的关键要素。杨勇等<sup>[9]</sup>引入了风险度向量方法, 对每个角色和任务的潜在风险进行量化评估, 从而更有效地管理和控制访问权限。同时, 为评估符合资质约束的用户群在执行多项任务时的聚集风险, 应用模糊综合评估法, 结合多个因素数据, 评估执行多任务时的总体风险。杨雪莱等<sup>[10]</sup>基于信号方法, 选取月度和季度预警指标, 逐一分析了中国金融风险预警指标的预警绩效, 确定了预警指标阈值。胡乐群<sup>[11]</sup>系统性地回顾了金融风险预警中的比较法、波动法和专家征询法等指标阈值的确定方法。

## 1.2 研究述评

目前, 企业技术风险属性衡量的研究主要基于产业经济理论和风险管理理论, 而从风险管理视角研究产业发展的算法模型的探索总体较少, 对于风险发生机制、阈值激活条件的理解仍有一定深入的空间。首先, 由于产业风险自身具有隐匿性、渐变性等诸多特性, 我国企业技术风险识别和监测的研究工作总体上处于早期发展阶段, 尤其是第二类风险预警研究往往围绕具体产业展开, 存在产业风险研究内在特征认识不足、成因机制了解不清、演化规律把握不清晰、解决方案不匹配等问题, 其方法模型将可部分适应性迁移至其他产业领域。其次, 国内学者对“风险识别”的研究多集中于宏观层面的体系描述性分析、定量研究, 常采用“风险因子指标体系梳理—根据专家意见或规则赋权—综合计算评分”的技术路线, 尚未形成统一框架。此类研究方法存在量化指标权重主观性强、数据获取难、风险预测前瞻性不足等问题, 个人或组织越来越需要从前瞻性、

系统性视角把握风险及其态势变化, 因此要逐渐加强对企业技术风险阈值方面的研究。

## 2 企业技术风险阈值激活模型设计

为使企业技术风险识别过程高效化、专业化且具备处理分散复杂的大数据能力, 为决策者提供风险控制依据, 本文将传统风险识别问题转化为利用机器学习进行二分类的预测问题。其工作方法涉及以下关键步骤: 对包括已标记的训练样本的数据集进行特征工程, 即选择和转换输入特征, 使模型能够更好地理解数据; 采用监督学习算法执行分类任务, 通过训练数据来学习模型的参数或规则; 在训练阶段中, 模型尝试优化目标函数以最佳拟合训练数据, 并从已知输入和输出关系中学习; 模型根据学习规则判断输入数据所属类别或预测目标值, 以预测未知企业是否存在技术风险。这种策略向企业提供了智能化的风险检测和预测工具, 确保了风险管理过程中决策的高效、科学和适应性。

### 2.1 研究流程设计

本文研究的目标在于通过明确定义企业风险并进行多维度的分类, 采用一种机器学习方法, 构建“企业技术风险阈值激活模型”(enterprise technical risk threshold activation model, ETRTA)以深度挖掘企业风险的特征。在研究过程中, 将运用包括Logistic回归等在内的10种不同算法来训练参数变量, 以学习企业技术风险属性并最终确定最佳的风险激活阈值, 并对模型的有效性进行全面评估。

基于机器学习的企业技术风险阈值激活流程如图1所示。

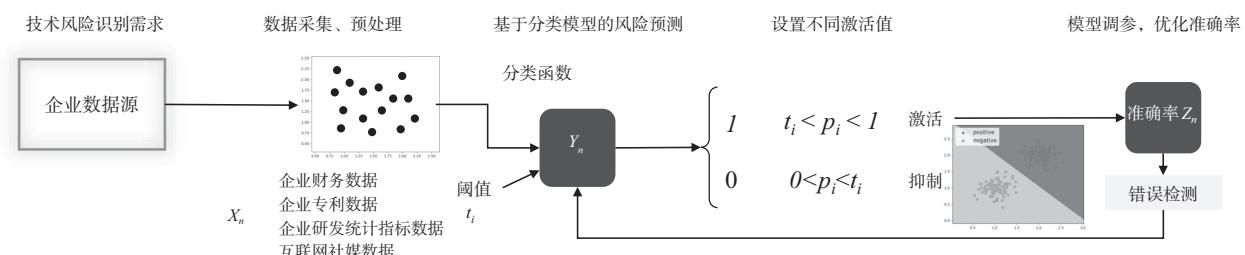


图1 研究流程及主要步骤

构建企业技术风险阈值激活模型关键是要明确所研究的技术风险类型边界，并根据技术风险因素需求从企业数据源中筛选与技术风险因素相关的数据。采用专业网络采集工具进行数据采集，以确保获取开源数据的全面信息。为保障数据的质量与可用性，执行系列数据预处理步骤包括数据清洗、特征筛选、数据标准与归一化、数据集划分以及不均衡数据处理，以完成模型输入数据的准备。在建模与训练迭代阶段，深入研究了集成学习方法，如随机森林、梯度提升树等多种机器学习模型，整合多个弱学习器以提升模型性能和稳健性。为了全面评估模型的泛化性能，精细划分数据集并进行多次训练和测试，以最小化过拟合的潜在风险。在建模过程中进行了阈值激活参数的调优，包括阈值等模型配置参数微调，以优化模型性能。最终研究专注于模型的评估与优化。模型训练完成后，运用多种评估指标对模型性能进行全面评估，以确保模型的高质量和可用性。

## 2.2 企业技术风险数据维度

企业技术风险定义为企业在从事技术创新、新技术应用、研发和生产等技术相关活动中所遭遇的潜在威胁和不确定性，可能对企业的技术进展、业务资产和市场竞争地位造成负面影响。企业技术风险涵盖了多方面来源，包括技术自身的持续演进、复杂非线性、依赖关联等内部风险源，以及市场不确定性、竞争压力、法律法规变化、知识产权、供应链中断、安全等外部风险源。在技术创新迭代过程中，企业受到先进技术不成熟、不配套以及对可靠性验证不足等制约。这种情况导致技术风险的成因和机制较为复杂，

技术不可预测性、替代技术缺乏等都可能导致创新失败。同时，技术本身的特性可能导致难以科学判断其市场适应性和先进性，这就意味着获取相关风险数据以及进行有效的识别和测度都面临一定困难。故本文未将因技术自身特性引发的风险纳入考虑范围。

表1所示的企业数据与外部技术风险相关描述项中，部分可以直接反映技术开发的经费投入强度，包括研发预算、人员配置以及研发项目的进展情况。在知识产权方面，专利、商标和版权数据通常被用来评估企业的技术创新能力和知识产权保护。然而，有关技术人才的数据获取可能较为困难，涉及其技术背景、能力以及参与技术创新项目的信息。此外，项目成功率、产品开发周期和技术需求等数据指标更多地依赖于主观评估和专家知识，难以反映整体趋势。

综合分析企业技术创新风险识别的主要数据来源，考虑研究领域特征、数据代表性、获取便利性以及模型各维度区分度等因素，将技术风险的识别边界限定在企业的经营状况、研发活动以及知识产权获得情况等外部数据。以上数据用于预测可能出现的企业技术风险更具实际可行性。

其中，财务状况与技术投入密切相关。企业的财务健康通常能够直接提高技术投入水平，因为充足的财务资源有助于研发和创新，进而降低技术风险。相反，财务困境可能会限制研发预算，增加技术风险。研发活动的影响在于，高强度的研发通常与更多的技术风险关联，因为创新尝试和新技术引入可能伴随失败。因此，在竞争激烈的技术密集型产业中，企业通常通过承担技术不确定性风险换取竞争优势机会。技术合作、

表1 常用企业数据中和技术风险相关的描述项

序号	描述项分类	数据指标
1	基于企业内部技术、生产、财务、市场等面板统计数据 <sup>[12]</sup>	财务报表、研发投入、劳动生产率、专利产出 <sup>[13-14]</sup> 、产品销售数据 <sup>[1]</sup> （利润率、市场占有率等）、净现值率、内部收益率、R&D经费占企业销售额比重、技术开发人员比重、获得风险资本数量等 <sup>[15]</sup>
2	基于多源异构文本非结构化信息 <sup>[1]</sup>	企业年报 <sup>[16]</sup> 、招股书 <sup>[17]</sup> 、临时公告、项目报告、媒体报道、产品评论中有关企业经营状况、发展动态、研发项目等历史披露数据 <sup>[1]</sup>
3	基于专家知识、问卷调查的分析预测数据 <sup>[18]</sup>	项目成功率、专利获得量、产品开发周期、开发资金回收周期 <sup>[15]</sup> 、技术需求分析、技术性能分析、技术改进分析、技术迁移分析、自主技术在产品开发中占的比重、行业技术进步速度等

专利许可和转让等活动是评估企业技术创新能力的重要指标。持有关键专利的企业表明其在技术领域的实力，并易在市场上形成竞争壁垒。专利的有效性和普及度也会影响企业的竞争地位。相比之下，竞争对手拥有强大的专利组合可能削弱本企业的技术竞争力。在此种情况下，企业可能会面临支付高额专利费用、市场进入风险乃至技术淘汰的风险。

### 3 智能网联汽车企业技术风险实证研究

#### 3.1 数据收集与处理

智能网联汽车是指搭载先进的车载传感器、控制器、执行器等装置，并融合现代通信与网络技术，实现车与X（车、路、人、云等）智能信息交换、共享，具备复杂环境感知、智能决策、协同控制等功能的新一代汽车<sup>[19]</sup>。此产业覆盖汽车、通信、互联网等多个领域，引发了国际新一轮科技竞争。对智能网联汽车企业技术风险进行分析，能够为我国汽车技术布局及相关产业和价值链体系的重塑提供参考。

鉴于此，本文选择中国4238家与智能网联汽车产业相关的上市和非上市企业（其中有176家企业出现技术风险）为研究对象，样本数据包括2017年12月31日—2022年12月31日的企业财务报告、研发活动、专利信息以及技术进行风险相关新闻报道。前3项样本所用数据全部来自CSMAR中国经济金融研究数据库和Datago财经研究数据库，第4项样本数据来自于网络大数据采集，经统一整合后形成企业风险数据集。数

据处理与分析软件采用Python 3.9.7。

##### 3.1.1 数据集描述

数据集包括企业财务、研发、专利等多源数据。数据元素的具体信息如表2所示。

从表2可以看出，样本中的财务指标、知识产权、研发投入数据自变量维度共计17项，记为 $X_i$ ；因变量Y为根据企业在新闻报道提及技术风险事件情况，包含知识产权、技术竞争、研发创新、技术替代风险等类型，构建企业风险分类标签。设置one-hot变量以反映企业在过去是否发生风险事件，将模型预测同企业现实表现相比对。具体设定方式为：对于8个模型计算出的风险发生概率，正样本表示期望模型识别或预测的事件，则如果企业发生风险，相应的默认Y值设为1，若无风险则设为0。

##### 3.1.2 数据预处理

为了提高数据的质量和可用性，采用了数据清洗、特征筛选及数据标准与归一化等预处理方法。在4238条源自互联网公开信息的样本数据中，222条样本存在缺失值，占总样本的5.23%。考虑到缺失值数量较少，因此选择剔除这些数据，最终保留4016条有效样本。由于不同维度数据的量纲不一致，对17维 $X$ 数据执行了Z-score标准化处理，以实现数据缩放。在4016条有效样本中，正样本仅有176条，即数据集中仅有176家企业面临技术风险，占比为4.43%，呈现出样本分布的明显不均衡性。为防止模型训练偏差，对数据进行了均衡处理。通过对负样本进行10次随机欠采样，生成10个样本子集，每个

表2 数据元素集

变量	数据维度	数据来源	数据元字段
自变量 $X_1, \dots, X_8$	财务数据	公司主要财务指标表	总资产、总负债、无形资产、归属于公司股东的净利润、净利润、营业收入、营业成本、营业利润
自变量 $X_9, \dots, X_{15}$	研发数据	研发投入情况表	研发人员数量、研发人员数量占比（%）、研发投入金额、研发投入占营业收入比例（%）、研发投入（支出）费用化金额、研发投入（支出）资本化金额、资本化研发投入（支出）占研发投入比例（%）
自变量 $X_{16}, X_{17}$	专利数据	国内外专利申请获得情况表	申请专利数量、申请发明专利数量
预测变量Y	技术风险数据	新闻基本信息表	新闻标题、新闻全文、是否提及技术风险

注：数据区间为2017-12-31—2022-12-31。

子集中负样本的采样数为 384 条, 与包含 176 条记录的正样本集组合后配对形成数据集, 将训练集和测试集按照 6:4 比例分割。

### 3.1.3 数据集特征分析

为了评估变量间的关系以及确定这些关系是否具有统计学意义, 对处理后的数据进行相关性与显著性分析, 如图 2 所示。

分析表明, 在 17 项自变量中, 部分指标如研发人员数量、研发投入(支出)资本化金额等相关系数检验概率  $p$  值近似为 0。表明当显著性水平为 0.05 时, 拒绝相关系数检验零假设, 两者总体存在线性关系, 考虑到双方 Pearson 相关系数没有超过 0.300, 因此不具备强相关性。对于其他概率  $p$  值较大指标, 当显著性水平为 0.05 时, 应接受相关系数检验的零假设, 两者总体不存在线性关系。

## 3.2 模型构建

在综合考虑本文研究的数据集规模、数据

完整性、静态性以及变量特征等多方因素后, 采用了一种人工标注风险训练集的方法进行迭代训练。具体选用了 8 种机器学习模型, 包括 Logistic 回归、支持向量机、K 临近、朴素贝叶斯、随机森林、梯度提升树、XGBoost 和 LightGBM。这 8 个模型的优劣势对比如表 3 所示。

将每个企业的全部参数分别作为特征输入模型, 即找出  $Y=f(X_i)$  的函数关系, 从而根据  $X_i$  值预测出  $Y$  值。这些模型用于对每个输入数据进行二分类预测, 共进行了 80 (10 个数据集  $\times$  8 个模型数量) 次模型训练, 全面评估这些模型在不同数据集和情境下的性能。

## 3.3 模型评估

本文主要采用准确率、精确率、召回率、ROC 曲线中的  $AUC$  值和混淆矩阵等指标全面评估算法模型的适用性及效果。

对过去 10 年内的企业数据测试结果, 绘制不同模型的准确率热力图矩阵, 高准确性地表明

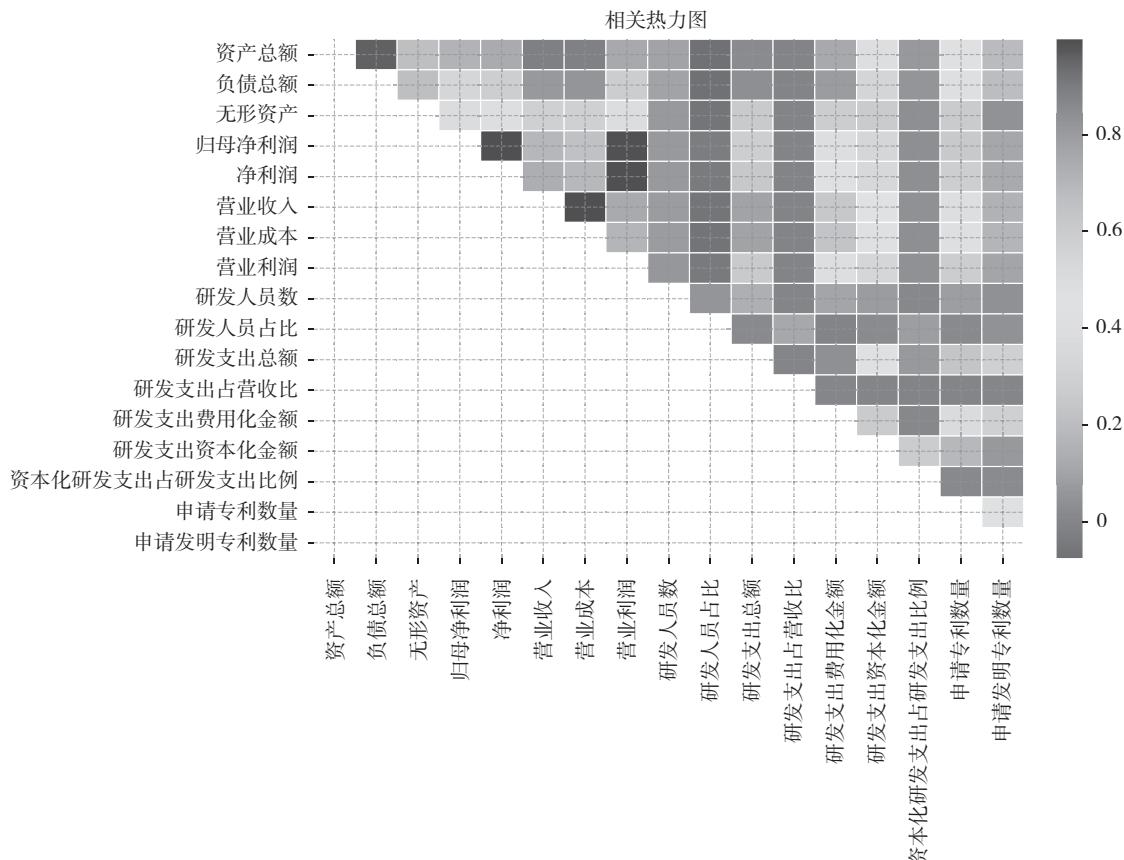


图 2 相关性与显著性分析

其在识别该领域技术风险方面具有一定优势。根据图 3a 不同算法的预测结果可知, 基于验证集或测试集的计算结果, 随机森林及其他 3 种梯度提升模型的准确率均在 80% 以上, 预测表现良好。使用 LightGBM、GBDT、XGBoost 模型的综合模型, 其平均模型准确率最高, 达到 82.59%。

通过常用的分类准确程度的统计方法——ROC 曲线定量比较 8 个模型的风险预测能力。本文 ROC 曲线的绘制方法为: 数据根据模型预测正类的概率对数据进行排序。对于每个可能的阈值计算对应的真正率 ( $TPR$ ) 和假正率 ( $FPR$ )。以假正率 ( $FPR$ ) 为横轴, 真正率 ( $TPR$ ) 为纵轴, 绘制 ROC 曲线。ROC 曲线与 45 度线之间所围图形的面积即为  $AUC$  值。观察图 3 右侧的 ROC 曲线可视化结果可知, 模型整体的分类性能出现一定程度的差异。而单一模型的  $AUC$  值情况不一, 除 MLP 及 KNN 外, 其他 6 种算法的  $AUC$  值均超

0.7, 表现出较高的准确性。从图 3 可知, 综合模型的 ROC 曲线呈现出色性能。

根据表 4 混淆矩阵中的实际预测结果, 计算精确率可知, 模型在所有预测为正例的样本中, 有 88.89% 是真正例。而模型召回率较低, 仅能够识别出 54.05% 真正例, 表明模型在识别正例方面存在改进空间, 以降低风险漏报率。

在风险预测场景下, 阈值激活模型使用基于综合模型的分类方程来表示。每个风险指标的预测概率都对应一个阈值, 当预测概率超过阈值时, 阈值激活函数产生输出。激活阈值通常由模型计算风险未来发生的预测概率  $p$  来确定, 相对于风险波动性, 风险阈值激活的基线水平通常设定为 0.5。对于二分类阈值激活器, 当风险指标值  $p >$  阈值  $t$  时, 阈值激活函数的输出为 1, 否则输出为 0。图 4 横向比较了不同阈值下的模型准确率, 发现当阈值设置为 0.46 和 0.52 时, 模型

表 3 机器学习模型横向对比

模型	模型计算难度	处理问题类型严格程度	适用数据维度	样本规模限制	对异常值敏感程度	易过拟合程度
Logistic 回归	低	线性	低维	小	低	高
支持向量机	高	非线性	高维	大	高	高
K 临近	高	非线性	高维	大	高	高
朴素贝叶斯	低	线性	高维	小	低	低
随机森林	高	非线性	高维	大	低	高
梯度提升树	高	非线性	高维	大	高	高
XGBoost	高	非线性	高维	大	低	高
LightGBM	低	线性	高维	小	高	低

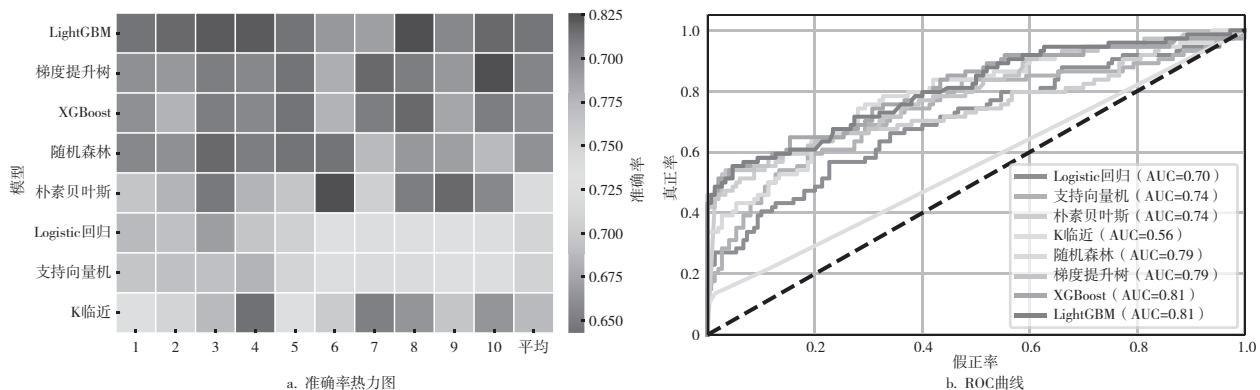


图 3 模型准确率及 ROC 曲线

表 4 0\_9 样本混淆矩阵

样本数量	实际为 0	实际为 1
预测为 0	145	34
预测为 1	5	40

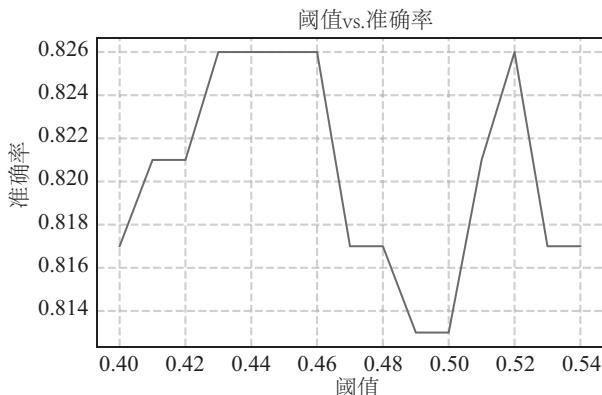


图4 模型在不同准确率下的激活阈值

准确率达到极大值。对于实际存在风险的企业，本文倾向于提高被正确预测的比例，即提高召回率。故当阈值设置为0.46时，模型更倾向于将样本预测为1，此时模型更为严格。

#### 3.4 模型输出结果分析

随着汽车产业朝着电动化、网联化、智能化方向迅速演变，智能网联汽车产业正处于技术创新、产业转型和产品升级的关键时期。对于部分存在风险且被模型成功预测的企业，形成了包含以下5个维度的技术风险画像，这些维度反映了企业面临的不同技术挑战类型。一是新兴技术在业务域应用效果的不确定性。智能网联汽车在政务、民生、金融、智能制造等领域具备大量应用场景。若市场需求不及预期，可能面临着产能过剩和市场错判风险。二是核心元器件、零部件供应风险。智能网联汽车数据处理量巨大，对汽车芯片要求极高，随着智能汽车E/E架构升级、域控制器/中央计算平台广泛使用，车载AI芯片市场规模将持续增长。但当前全球行业普遍存在芯片短缺问题，可能导致制造商减产或停产，对企业的产能和毛利率水平构成直接威胁。长远来看，整个产业链上的核心技术，如车载芯片和操作系统，都可能面临“卡脖子”技术风险。三是面对新技术冲击时的技术更新迭代不及时的风险。智能网联汽车融合感知、决策、通讯、执行、智能驾驶等众多前沿技术，若企业未能及时调整产品架构或对技术策略进行有效布局以适应或引领技术进步，可能面临与行业变革速度不同

步的技术发展风险，导致企业在技术竞争中处于劣势。四是技术创新与竞争风险。随着下游行业变革，许多传统的交互、控制和显示技术已经变得不太适用。为了提升产品中技术含量，企业应积极加强研发，并在专利方面进行策略性布局。

数据选择是本文研究模型适用性的关键因素。在数据收集过程中，考虑了特定行业各种不同类型和规模的企业，包括初创公司、中型企业、大型企业和国有企业。因此，这个模型不仅适用于当前研究的智能网联汽车行业，也适用于其他战略性新兴和技术密集型行业，如生物技术、可再生能源等领域。鉴于智能网联汽车属于发展中市场，本模型的适用性可能受限于市场形态导致的数据分布差异，且还受限于其他领域因数据不透明或市场监管不健全导致的数据可获得性和质量差异。

本文研究的模型设计具有高度灵活性，可根据不同行业的特定需求进行优化，模型中的风险评估参数可以根据行业特定的风险元素进行定制化设置，如在医疗行业中加入临床试验作为考量因素。此外，这个模型还可以与其他分析技术如强化学习、智能体仿真等结合使用，以进一步提高预测的准确性和效率。这种扩展性让ETRTA模型不仅适用于目前的产业研究，还能够适应未来的技术趋势分析和风险管理。

## 4 结论与展望

本文构建了一种风险识别半自动机器辅助模型。这一模型利用自动化方法以量化分析产业风险激活的规律性特征，实现从财务、研发、专利、新闻相关数据变量中识别出有潜在技术风险的企业。在构建的分类模型中，基于LightGBM、GBDT、XGBoost三者综合模型分类预测准确性达到82.59%，说明模型可较好地识别出技术风险企业，但决策部门需要结合实际情景权衡漏报率与命中率代价。模型训练变量提供了对企业发展产生积极或消极影响的技术风险因素，在对样本的预测效果上呈现一定的显著性。可支持发现与企业风险相关性较高的风险变量，帮助决策者了解外部环境中企业风险变化。0.46的阈值参数为

合理划设企业风险的阈值区间提供借鉴, 有效地避免了因阈值划设过松或过严而出现的漏警及虚警的问题, 对于风险弱信号识别的效率和准确性具有操作意义。

然而, 模型准确性仍面临多方面的限制。首先, 模型依赖数据源自公开渠道, 这可能导致在某些特殊领域的企业技术风险评估中, 由于数据不足影响预测结果的准确性。此外, 模型所利用的数据维度有限, 这种信息的不完整性可能限制了对风险因素的全面考量。未来的研究方向应着重于扩展数据源和指标, 考虑不同领域对各维度数据的认可度差异, 以构建更加综合的评价维度, 提升模型成功拟合实际影响因素的概率。同时, 为缓解数据量对模型性能的制约, 计划采集更多样本数据进行深入训练, 进而同步提升模型的预测准确性和稳定性。在此基础上, 提出探索领域迁移的可能性, 通过调整训练集的对标对象, 将模型优先应用于其他相近领域的企业技术风险评估工作中, 从而扩大模型适用范围, 如燃料电池汽车、氢能源汽车等领域。这些研究方向有望进一步提高模型的应用潜力和实际效用。

## 参考文献

- [1] 宋姗姗, 钟永恒, 刘佳, 等. 企业技术创新风险研究进展与热点议题[J]. 技术与创新管理, 2023, 44(5): 525–540.
- [2] 郝玲玲. 技术创新风险的哲学思考: 银星创新始末研究[J]. 文化创新比较研究, 2019, 33(3): 31–32.
- [3] 朱爱孔, 谷佳璐, 魏美娟, 等. 危机管理视角下新疆棉花产业风险应对策略研究[J]. 中国棉花加工, 2022(4): 4–9.
- [4] 雷志梅. 基于知识元的产业经济风险扩散复杂网络模型研究[D]. 大连: 大连理工大学, 2018.
- [5] 杨君良. 产业共生网络的风险传播机制及缓解策略研究[D]. 南京: 南京信息工程大学, 2021.
- [6] 林金枫. 我国新能源汽车产业风险识别与评价研究[D]. 哈尔滨: 哈尔滨工程大学, 2015.
- [7] 王粟旸, 肖斌卿, 周小超. 外部冲击视角下中国银行业和房地产业风险传染性测度[J]. 管理学报, 2012, 9(7): 968–974.
- [8] 夏喆. 论企业财务风险的传导机理[J]. 财会月刊, 2009, 29(7): 19–22.
- [9] 杨勇, 孙宇清, 周丽. FSP: 一种基于风险的安全策略[J]. 计算机工程与应用, 2010, 46(13): 82–86.
- [10] 杨雪莱, 许传华. 中国金融风险预警指标的最优阈值及预测绩效分析[J]. 广东金融学院学报, 2012, 27(2): 121–128.
- [11] 胡乐群. 风险预警中指标阈值确定方法[J]. 金融电子化, 2011(9): 43–45.
- [12] MU J F, PENG G, MACLACHLAN D L. Effect of risk management strategy on NPD performance[J]. Technovation, 2009, 29(3): 170–180.
- [13] GE X Y. Key element identification in cooperative technological innovation risk on social network analysis[C]//Proceedings of the 2014 Seventh International Joint Conference on Computational Sciences and Optimization, July 4–6, 2014, Beijing University of Chemical Technology, Beijing. Washington: IEEE Computer Society, 2014: 316–319.
- [14] 李莹, 周胜生. 基于大数据视角的企业专利风险预警研究[J]. 情报杂志, 2021, 40(10): 21–26.
- [15] 宗明华. 企业技术风险及其分析[J]. 昆明理工大学学报: 理工版, 2001, 26(1): 118–122.
- [16] 梁娜, 姚长青, 王峰, 等. 基于三重维度的企业风险信息抽取方法研究[J]. 情报学报, 2019, 38(12): 1241–1249.
- [17] 饶先成. 科创板拟上市企业的专利风险识别及应对: 基于205份科创板拟上市企业招股说明书的实证分析[J]. 科技与法律, 2020(6): 31–40.
- [18] MASÁR M, HUDÁKOVÁ M. Project risk identification of a new product in enterprise in Slovakia[C]//CBU International Conference Proceedings 2018, March 21–23, 2018, Central Bohemia University, Prague. Prague: CBU Research Institute, 2018: 343–349.
- [19] 国家制造强国建设战略咨询委员会. 《中国制造2025》重点领域技术路线图(2015版)[M]. 北京: 国家制造强国建设战略咨询委员会, 2015: 9–15.