

企业的数字化转型如何发展？

——基于机器学习的证据

林泓兵, 张浩*, 张志强

南京林业大学经济管理学院, 江苏 南京

收稿日期: 2025年11月7日; 录用日期: 2025年11月24日; 发布日期: 2025年12月29日

摘要

数字化转型是推动制造业高质量发展的关键路径,对提升企业资源配置效率与核心竞争力具有深远影响。多数企业虽然认识到数字化转型的重要性,但对企业如何发展数字化,其关键影响因素是什么仍模糊不清。本文基于2015~2024年中国A股上市公司数据,采用XGBoost等机器学习方法,系统识别企业数字化转型的关键影响因素,并分析其非线性影响机制。研究发现:技术人员占比、行业竞争强度与专利数量是影响数字化转型最为重要的三大因素,且均存在显著的非线性特征。具体而言,技术人员占比在突破30%阈值后呈现边际效应递增;行业竞争强度表现为“倒U型”影响,拐点位于HHI = 0.25附近;专利数量则具有持续正向但边际递减的效应。此外,XGBoost模型在预测数字化转型方面显著优于传统回归方法。基于此,本文从强化技术人才储备、优化市场竞争环境、推动高质量专利转化等方面提出政策建议,为政府与企业推动数字化转型提供理论依据与实践参考。

关键词

数字化转型, 影响因素, 机器学习, XGBoost

How Can a Company's Digital Transformation Progress?

—Evidence Based on Machine Learning

Hongbing Lin, Hao Zhang*, Zhiqiang Zhang

College of Economics and Management, Nanjing Forestry University, Nanjing Jiangsu

Received: November 7, 2025; accepted: November 24, 2025; published: December 29, 2025

*通讯作者。

文章引用: 林泓兵, 张浩, 张志强. 企业的数字化转型如何发展? [J]. 电子商务评论, 2025, 14(12): 4981-4992.
DOI: 10.12677/eci.2025.14124452

Abstract

Digital transformation is a key path to promoting high-quality development of the manufacturing industry, which has a profound impact on improving the efficiency of enterprise resource allocation and core competitiveness. Although most companies recognize the importance of digital transformation, the key influencing factors on how they develop digitalization are still unclear. This article is based on data from Chinese A-share listed companies from 2015 to 2024, using machine learning methods such as XGBoost to systematically identify the key influencing factors of enterprise digital transformation and analyze their nonlinear impact mechanisms. Research has found that the proportion of technical personnel, industry competition intensity, and number of patents are the three most important factors affecting digital transformation, and all have significant nonlinear characteristics. Specifically, the proportion of technical personnel shows an increasing marginal effect after exceeding the 30% threshold; the intensity of industry competition shows an “inverted U-shaped” impact, with the inflection point located around $HHI = 0.25$; the number of patents has a sustained positive but marginally decreasing effect. In addition, the XGBoost model is significantly superior to traditional regression methods in predicting digital transformation. Based on this, this article proposes policy recommendations from strengthening the reserve of technical talents, optimizing the market competition environment, and promoting high-quality patent conversion, providing theoretical basis and practical reference for the government and enterprises to promote digital transformation.

Keywords

Digital Transformation, Influencing Factors, Machine Learning, XGBoost

Copyright © 2025 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

随着数字经济时代的到来，数字化转型已成为企业构筑核心竞争力、实现高质量发展的关键路径。无论是应对市场环境的动态变化，还是满足消费者日益个性化的需求，数字化能力都为企业优化资源配置、创新商业模式、提升全要素生产率提供了前所未有的机遇。当前，大多数企业虽然普遍认识到数字化转型的重要性，但在实践过程中仍存在方向不明、路径模糊的问题，尤其是对数字化转型的关键影响因素缺乏系统认知，难以判断哪些内外部因素在推动或制约转型进程中发挥核心作用。基于学习曲线与资源约束理论，企业在数字化技术人才积累过程中可能呈现规模效应，初期投入效果有限，而跨越某一临界点后效益显著提升；同时，依据产业组织理论，市场竞争程度可能与企业数字化创新投入呈倒U型关系——适度竞争激发创新，过度竞争则因资源分散与抑制效应相关。在这一宏观背景下，深入探究企业数字化转型的影响因素，对于推动实体经济与数字技术的深度融合、引导资本市场资源有效配置具有重要的理论与现实意义。

现有研究为理解数字化转型的动因提供了丰富的视角，主要可归纳为技术、组织与环境三大维度。在技术因素方面，学者们普遍认为数字技术(如物联网、云计算、大数据)的渗透与应用是转型的基础[1]，而企业的技术创新能力与研发投入则构成了转型的核心驱动力[2]。在组织因素层面，研究不仅强调了组织结构转型[3]与高管背景[4]的重要性，还深入探讨了数据要素[5]与企业家精神[6]在激发数字化潜能中

的关键作用。在环境因素方面,政府补助[7]、产业政策[8]及市场竞争等外部力量也被证实对企业数字化转型产生显著的推动或调节效应。

尽管现有成果丰硕,但仍存在以下局限:首先,多数研究依赖于传统线性回归模型,侧重于对单一因素“净效应”的因果识别,难以充分捕捉多变量间复杂的非线性交互作用。其次,在驱动因素的相对重要性排序上缺乏系统性比较,导致学界与实务界难以辨别哪些是“牵一发而动全身”的核心杠杆,哪些是辅助性条件。最后,虽有研究采用定性比较分析(QCA)等组态视角探讨多因素协同,但其结论在更大样本中的预测效度与稳健性仍有待于基于预测范式的模型加以验证。

为弥补上述研究不足,本文引入机器学习方法,对中国 A 股上市公司数字化转型的驱动机制进行探索性研究。具体而言,本文的研究思路如下:第一,以多元线性回归为基线模型,并系统对比 LASSO、决策树、随机森林与 XGBoost 等多种算法在预测企业数字化水平上的性能,以遴选出最优模型。第二,基于最优模型(XGBoost),运用 SHAP (SHapley Additive exPlanations)值系统评估各驱动因素的相对重要性,并借助 ALE (Accumulated Local Effects)图描绘关键因素的非线性影响路径与临界值,例如技术人才密度是否存在规模效应,以及市场竞争是否呈现倒 U 型关系。第三,通过替换被解释变量(吴非指数与赵宸宇指数)与机器学习算法(随机森林)进行稳健性检验,确保研究结论的可靠性。

本文的边际贡献在于:方法上,突破了传统线性模型的局限,利用机器学习优势更为精准地刻画了数字化转型驱动因素的复杂机制;理论上,不仅识别出技术、组织与环境维度中的核心变量,更基于学习曲线、资源约束与产业组织等理论,清晰地描绘了其非线性作用规律,为理解企业数字化转型的“黑箱”提供了新的经验证据;实践上,研究结论能为企业优化数字化资源配置、以及监管部门制定精准化的产业与人才政策提供数据驱动的决策依据。

2. 研究设计

2.1. 研究方法

本文旨在分析企业数字化水平的形成机制,并识别影响企业数字化发展的关键因素。与传统实证研究侧重因果关系不同,本研究属于预测性分析,侧重于比较不同影响因素对企业数字化水平的预测能力。基于此,本文采用机器学习方法进行探索性研究,同时将传统多元线性回归作为基线方法,以便对比分析预测性能与解释能力。

本文在研究中采用多种建模方法,以同时兼顾预测性能与解释能力。首先,多元线性回归(Multiple Linear Regression)作为传统统计方法,用于构建基线模型,并提供直观的解释性分析[9]。为了进行变量筛选并降低模型过拟合风险,本文引入 Lasso 回归(Least Absolute Shrinkage and Selection Operator),通过 L1 正则化压缩部分回归系数,将不重要的特征系数置零,从而实现模型简化[10]。此外,本文采用决策树(Decision Tree)模型,这是一种非参数方法,能够处理变量间复杂的非线性关系和交互效应,广泛应用于企业管理与财务预测研究[11]。为了进一步提升预测精度和模型稳定性,本文使用随机森林(Random Forest),该集成学习方法通过构建大量决策树并进行投票或平均,实现比单一模型更优的预测效果。最后,本文采用 XGBoost (Extreme Gradient Boosting)算法。该算法是梯度提升树(GBDT)的高效实现,具有出色的非线性建模能力和高维数据处理能力,同时通过正则化技术降低过拟合风险。

在超参数调优过程中,本研究采用 GridSearchCV 结合 5 折交叉验证的方法,在预设参数空间内搜索最优组合。XGBoost 作为本研究的核心模型,其最终确定的关键参数包括: $n_estimators = 200$ 、 $learning_rate = 0.1$ 、 $max_depth = 7$ 、 $subsample = 0.8$ 、 $colsample_bytree = 1.0$ 。经调优后的 XGBoost 模型在预测精度与模型稳定性方面均优于其他模型,是研究中性能最优的算法,因此被作为后续解释性分析(如 SHAP 值分

解)的主要模型(详细验证见 4.1)。

2.2. 模型性能评估方法

在评估模型性能时, 本文从解释能力和预测能力进行考量。模型解释能力反映了模型在预测企业数字化水平时的拟合程度。具体而言, 本文使用样本内拟合优度 R^2_{in} 评估模型对训练集样本的拟合效果, R^2_{in} 决定系数越高, 样本内拟合程度越高, 模型训练效果越好。为衡量模型在新样本上的泛化能力, 本文进一步采用样本外拟合优度 R^2_{oos} 和样本外可解释方差 EVS_{oos} 评估模型在测试集上的预测效果, 其值越大, 说明模型对企业数字化水平的预测能力越强, 样本外拟合效果也越好。通过这三项指标, 本文综合考察不同方法在训练与测试阶段的拟合和预测表现, 从而优选出最适合企业数字化研究的算法。

3. 数据来源和变量定义

3.1. 数据来源

本文选择 2015~2024 年 A 股上市公司作为起始样本, 在剔除 ST 及 ST*和企业数据缺失的样本后, 最终保留了 17,698 个观测值。选取 2015 年为样本起点是考虑到数字化转型这一概念是在 2015 年提出。本文的数字化转型数据使用吴非测度的企业数字化转型程度, 企业财务治理数据等其他数据来自国泰安数据库(CSMAR)和统计年鉴, ESG 表现数据来自华证 ESG 评级体系。为了降低异常值对数据分析的干扰, 本文对处于最顶部 1%和最底部 1%的数据点进行了缩尾处理, 并在训练算法时对数据进行标准化处理。

3.2. 变量定义

本研究的变量分为两类(见表 1): 被预测变量和预测变量。其中: 被预测变量为企业数字化转型程度; 预测变量为企业数字化转型的影响因素。变量选取的逻辑如下:

被预测变量, 数字化转型程度(Dig)。本文借鉴吴非的方法, 基于上市公司年报文本分析进行构建。该方法利用 Python 和 Java PDFbox 库提取年报全文, 并通过一个涵盖“底层技术运用”(人工智能、大数据、云计算、区块链)与“技术实践应用”(移动互联网、智能化应用、数字金融等)的综合性关键词库进行文本匹配, 统计关键词出现的总频数。为消除分布偏态, 最终对总词频加 1 后取自然对数, 作为企业数字化转型程度的代理变量。该指标数值越大, 表明企业数字化转型程度越高[12]。

预测变量, 数字化转型的影响因素。企业数字化转型是一个多层级驱动的系统性过程: 在技术层面, 数字化赋能通过渗透生产经营流程, 驱动资源配置优化与决策机制智能化; 在组织层面, 适应性重构通过调整治理结构与协同模式, 激发内部资源的整合与创新效能; 在环境层面, 外部生态则通过行业竞争态势与宏观经济政策, 共同构成转型的外生驱动与约束条件。基于这一逻辑, 本文将数字化转型的影响因素划分为技术维度、组织维度和环境维度三个方面, 并选取如下指标:

技术能力是企业数字化发展的核心基础, 主要体现为研发投入强度、技术人员占比、无形资产占比及专利数量等指标。这些指标能够反映企业在研发与创新方面的资源投入和技术储备, 从而支持数字化技术的应用与推广。

组织资源与治理结构影响数字化转型的执行效率和内部协同效果。本研究选取企业规模、资产负债率、净资产收益率、两职合一以及股权集中度等指标, 以反映企业治理能力、财务稳健性及决策效率。这些因素能够影响数字化转型的组织基础和资源整合能力。

外部环境为企业数字化发展提供制度约束和市场激励。本研究选取市场竞争程度、融资约束和地区国内生产总值(GDP)等指标, 反映行业竞争压力、资金可获得性及宏观经济发展水平。这些因素通过外部

压力和政策引导对企业数字化进程产生显著影响。

Table 1. Variable definition and description
表 1. 变量定义与说明

变量性质	变量名称	变量符号	变量测度
被预测变量	数字化转型程度	Dig	年报中与数字化转型相关细分指标词频的对数
	研发投入强度	RD	研发投入/营业收入
	技术人员占比	Tech	专业技术职称人员人数/总人数
	无形资产占比	Int	无形资产/总资产
	专利数量	Patent	专利获得总数加 1 取自然对数
	企业规模	Size	公司总资产的自然对数
预测变量	资产负债率	Lev	总负债/总资产
	净资产收益率	ROE	企业净利润/股东权益
	两职合一	Dual	董事长与总经理是同一个人 为 1，否则为 0
	股权集中	Top3	前三股东持股数量/总股数
	上市年限	Age	企业上市年限
	市场竞争程度	HHI	赫芬达尔指数
	融资约束	SA	SA 指数
	国内生产总值	GDP	各省份 GDP

4. 实证结果与分析

4.1. 不同机器学习算法的预测效果

根据表 2 可见，不同算法在样本内与样本外拟合效果上存在显著差异。传统统计方法(多元线性回归与 Lasso 回归)的样本内拟合优度 R^2_{is} 和样本外拟合优度 R^2_{oos} 均约为 0.20，表明其在捕捉企业数字化水平特征方面的解释能力有限。相比之下，基于非线性结构的机器学习算法表现出更优的预测性能。决策树算法的 R^2_{is} 达到 0.4546，说明其在训练样本上具备较强的拟合能力，但其样本外拟合优度 $R^2_{oos} = 0.3450$ 略低，显示出一定的过拟合倾向。随机森林算法通过集成多棵决策树，有效缓解了单一模型的偏差问题，样本外拟合优度提升至 0.3828，可解释方差(EVS_{oos})为 0.3829，说明其在泛化能力上优于前述模型。而综合来看，XGBoost 算法在所有算法中表现最优，其 $R^2_{is} = 0.8628$ ， $R^2_{oos} = 0.5648$ ，可解释方差(EVS_{oos})亦达 0.5649，

Table 2. The predictive performance of different machine learning methods
表 2. 不同机器学习算法的预测效果

预测方法	R^2_{is}	R^2_{oos}	EVS_{oos}
多元线性回归	0.2027	0.2108	0.2108
Lasso 回归	0.2027	0.2107	0.2108
决策树	0.4546	0.3450	0.3450
随机森林	0.4271	0.3828	0.3829
XGBoost	0.8628	0.5648	0.5649

显著高于其他方法。这表明 XGBoost 算法能够在充分学习训练样本特征的同时,保持良好的样本外预测能力。该算法在梯度提升框架下融合了树模型的非线性拟合优势与正则化约束机制,兼顾模型复杂度与稳健性。因此,本文认为 XGBoost 在预测企业数字化水平方面具有最佳综合表现,是最优的方法选择。

4.2. 不同因素对数字化转型程度的影响分析

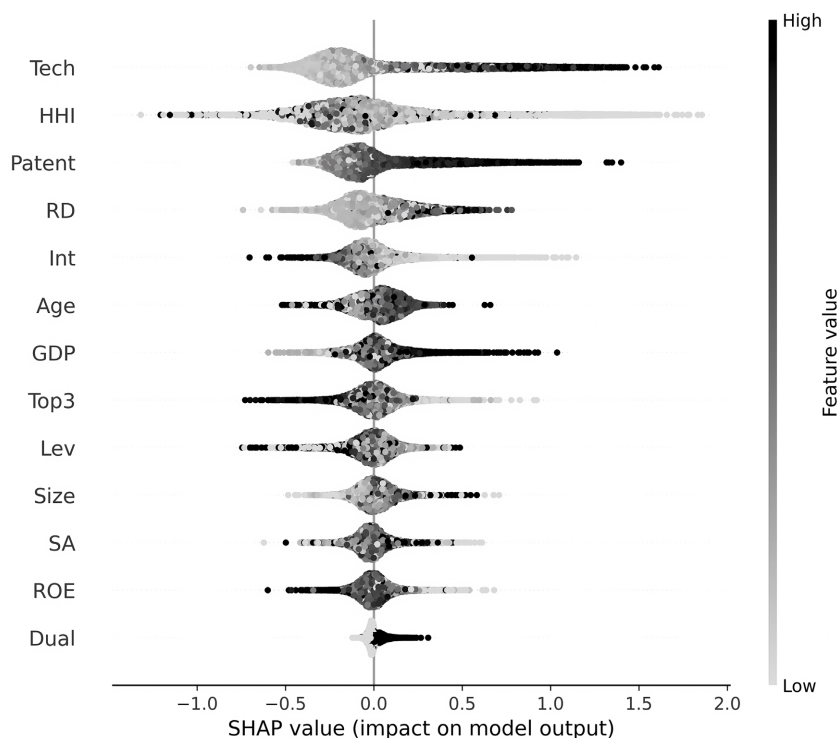


Figure 1. Summary plot of influencing factors (XGBoost)

图 1. 影响因素摘要图(XGBoost)

根据图 1 中 SHAP 值的分析结果,可以系统揭示各因素对企业数字化转型程度的影响方向与相对重要性。整体来看,变量对模型输出(即企业数字化水平)的边际贡献存在明显差异,表明企业数字化进程受多维因素共同驱动。

从影响强度来看,技术人员占比(Tech)、行业竞争程度(HHI)和专利数量(Patent)的 SHAP 值绝对值较大,说明其对数字化水平的解释力最强。其中,技术人员占比(Tech)的高取值对应明显的正向 SHAP 值分布,说明技术人力资本越丰富的企业,其数字化水平越高。这一结果体现了技术能力在数字化过程中的核心地位。行业竞争程度(HHI)亦呈现类似趋势,高竞争市场环境(颜色较深)主要分布在正向 SHAP 区间,表明外部竞争压力与企业数字化水平提升存在正相关关系。

其次,专利数量(Patent)与研发投入强度(RD)均表现出高值正向集中的特征,说明创新产出与研发活动能有效推动企业数字化进程。这表明企业技术创新活动不仅反映其知识积累水平,也能促进一定的数字化能力的形成。而无形资产占比(Int)的分布相对分散,但高值仍主要集中在正向区域,说明企业内部的知识与品牌资源能在一定程度上支持数字化投资。

在组织维度的变量中,规模(Size)的高取值对应正向 SHAP 值,说明大型企业因具备更强的资金与技术资源,更容易推进数字化转型。资产负债率(Lev)则呈现明显的双向影响:高值既可能增强数字化投入能力(通过杠杆效应),也可能因债务负担抑制数字化进程,反映其在不同财务约束条件下的非线性作用。

相对而言,股权集中度(Top3)的深色点部分分布在负向区域,暗示高集中度可能带来治理僵化,不利于创新性数字化决策。两职合一(Dual)的影响较弱,但其高值主要集中在 SHAP 值接近零或略为负向的位置,表明权力集中并未显著提升数字化水平,甚至可能因决策路径依赖而抑制数字化改革。

4.3. 关键驱动因素对企业数字化转型的预测模式

由上一节的研究结论可得, SHAP 值排名前三的“技术人员占比”、“行业竞争强度”和“专利数量”是数字化转型的关键影响因素。进一步地,本文借助累计局部效应图(Accumulated Local Effects Plot, ALE)能够排除了特征间相关性影响,侧重于探讨单一特征对模型预测作用的特点,继续分析关键影响因素对企业数字化转型的预测模式。累计局部效应图的横坐标为具体某个影响因素的特征值,纵坐标为影响因素的 ALE 值,该值大于 0 说明影响因素对数字化转型具有促进作用,反之则为抑制作用,值越大说明影响越显著。

4.3.1. 技术人员占比(Tech)

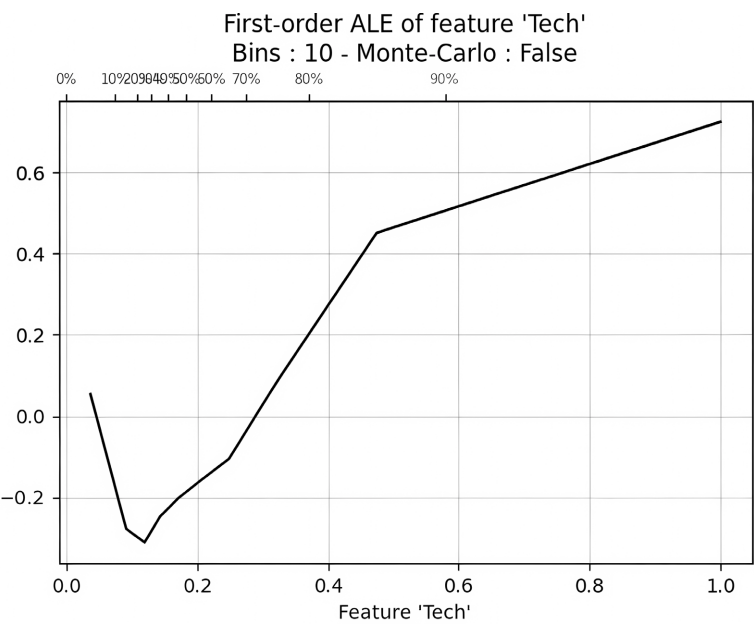


Figure 2. Technical staff proportion ALE chart
图 2. 技术人员占比 ALE 图

技术人员占比(Tech)对企业数字化转型程度的累积局部效应呈现出显著且稳健的正向单调递增关系。如图 2 所示,当技术人员占比处于较低水平(约 0~0.3 区间)时, ALE 值为负值且抑制效果在 0.1 时达到最大,表明此阶段技术人才基础的薄弱在一定程度上限制了数字化转型的效能释放。然而,当技术人员占比突破约 0.3 的临界点后, ALE 曲线呈现出急剧上升趋势,其边际效应显著增强。这表明,一旦企业技术人才储备积累到一定规模,将产生强大的协同效应与创新动能,极大地推动数字化转型的深度实施与迭代优化。随着技术人员占比进一步提升(超过 0.5), ALE 值的增长进入高位平台期,增速放缓但仍维持在高影响区间,反映出技术人才密度达到较高水平后,其对数字化的推动作用趋于稳定,持续为企业构筑坚实的数字化核心能力。

总体而言,该结论印证了技术人才资源是企业数字化转型的关键核心驱动力。其影响并非简单的线性关系,而存在一个关键的“规模效应阈值”。在达到该阈值之前,人才投入的回报有限;而一旦跨越该门槛,技术人才的价值将得到爆发式释放。这一发现从实证角度强调了,企业若想成功推进数字化转

型，需要优先进行前瞻性的技术人才布局与积累，以突破关键的人才瓶颈。

4.3.2. 行业竞争强度(HHI)

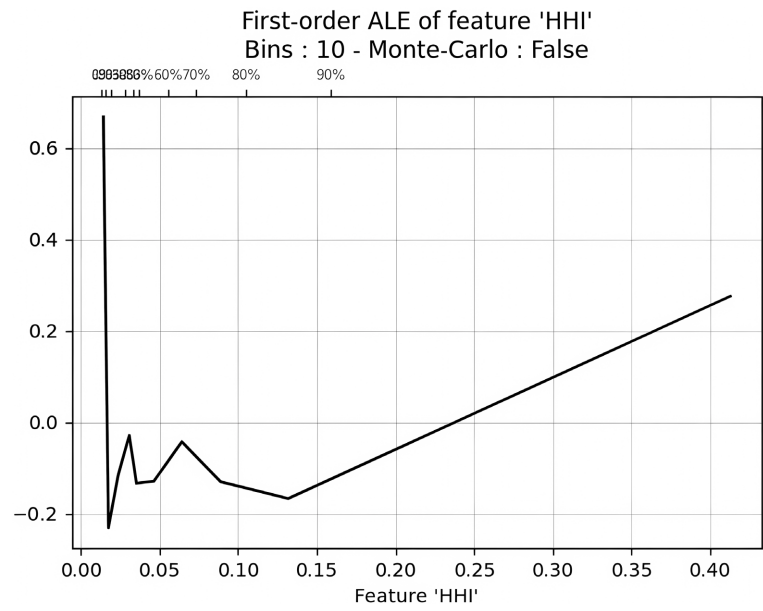


Figure 3. Industry competition intensity ALE chart
图 3. 行业竞争强度 ALE 图

图 3 中展示了行业竞争强度(HHI)对企业数字化转型程度的一阶累积局部效应变化趋势。整体来看，HHI 与数字化转型程度之间的关系呈现出典型的非线性正向效应特征。当 HHI 较低(约 0~0.1 区间)时，ALE 值波动较大且总体偏负，表明在竞争较为充分的市场环境中，企业面临较高的不确定性与成本压力，短期内可能抑制数字化投资的积极性。然而，当 HHI 超过 0.1 后，ALE 值迅速上升并在达到 0.25 时影响由负转正，说明随着市场逐渐集中、竞争强度适度上升，企业通过数字化转型提升管理效率与市场响应速度的动力增强，从而促进数字化水平提升。

总体而言，该结果表明行业竞争强度对企业数字化转型的影响具有“倒 U 型”的非线性特征：适度竞争能够强化企业的数字化激励，而过弱的竞争环境可能削弱数字化转型的积极性，从而阻碍企业的数字化进程。这一规律揭示了外部竞争环境在企业数字化转型中的约束与激励双重作用机制。

4.3.3. 专利数量(Patent)

由图 4 可知，专利数量(Patent)对企业数字化转型程度的累积局部效应呈现出持续正向但边际效应递减的特征。当专利数量处于低位(约 0~4 区间)时，ALE 值随专利数量增加而稳步上升，但整体仍为抑制效果，这表明创新成果的初始积累为企业数字化转型提供了技术基础与合法性，从而推动转型进程，但前期效果并不明显。当专利数量超过 3.5 以后，专利数量对数字化转型的促进效果开始显著，并且增速较快，直到专利数量达到 4.5 以后，ALE 曲线的上升斜率开始放缓。

这一模式揭示出，专利所代表的创新产出对数字化转型的驱动作用存在明显的效用边界。初期，专利作为技术创新能力的显性信号，并不能够有效支撑数字化战略的实施；但随着专利储备的持续增加，其对于数字化转型的贡献逐渐增加，并在达到一定阈值之后开始放缓，总体呈现边际效应递减。这表明，单一的专利数量积累并非推动数字化转型的持续动力，转型更依赖于组织对技术的消化、吸收和再创新能力，以及与之配套的战略、人才与流程变革。

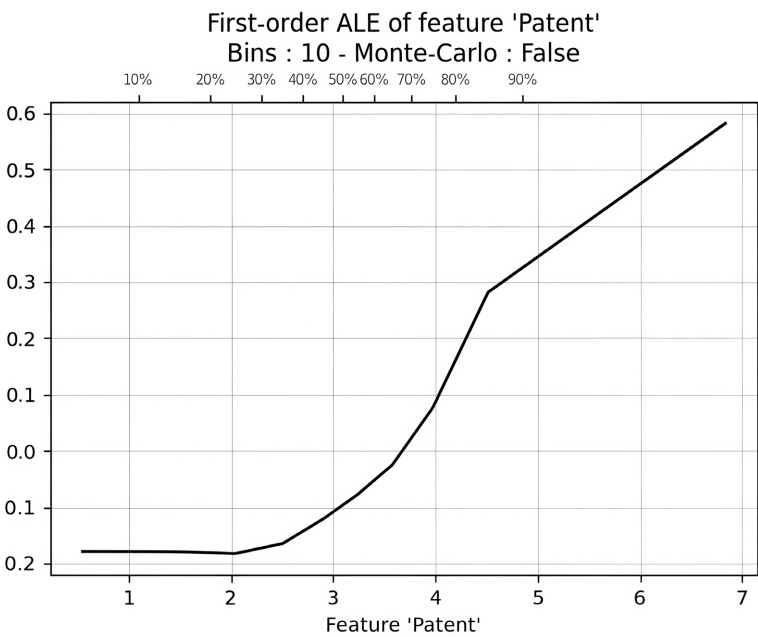


Figure 4. Number of patents ALE chart
图 4. 专利数量 ALE 图

结合技术人员占比、行业竞争强度等其他核心驱动因素的分析，专利数量的预测模式揭示了企业数字化转型的本质并非仅由技术成果堆砌而成，而是一个由人才基础、竞争环境与组织能力共同驱动的系统性变革过程。

4.4. 稳健性检验

4.4.1. 替换被预测变量

在实证分析部分，本文采用吴非计算的数字化转型水平作为被预测变量。同时，为进一步提升研究的普适性，本文采用赵宸宇等(2021)构建的数字化转型综合指数，作为替代性预测变量。该指标通过文本分析法提取企业年报中与数字化相关的特征词，并从技术应用、商业模式、智能制造及信息系统四个维度合成，数值越大代表企业数字化转型水平越高[13]。根据表 3 的结果，随机森林和 XGBoost 模型的拟合效果仍优于多元线性回归，研究结论保持一致。

Table 3. The predictive performance of different machine learning methods
表 3. 不同机器学习算法的预测效果

预测方法	R^2_{Is}	R^2_{oos}	EVS_{oos}
多元线性回归	0.1835	0.1937	0.1937
Lasso 回归	0.1764	0.1824	0.1825
决策树	0.5061	0.3305	0.3305
随机森林	0.5058	0.4219	0.4219
XGBoost	0.9263	0.5965	0.5966

4.4.2. 替换机器学习算法

为检验研究结论的稳健性，本文替换了机器学习算法进行对比分析。将主模型 XGBoost 替换为随机

森林重新进行训练与 SHAP 分析。结果显示(见图 5), 随机森林模型中特征的重要性排序(Tech 与 HHI 最为重要)及其对数字化转型的边际影响方向, 与主模型结论高度一致。这表明本文关于数字化转型核心驱动因素的发现, 对于不同的模型设定具有稳健性, 并非特定算法下的偶然结果。

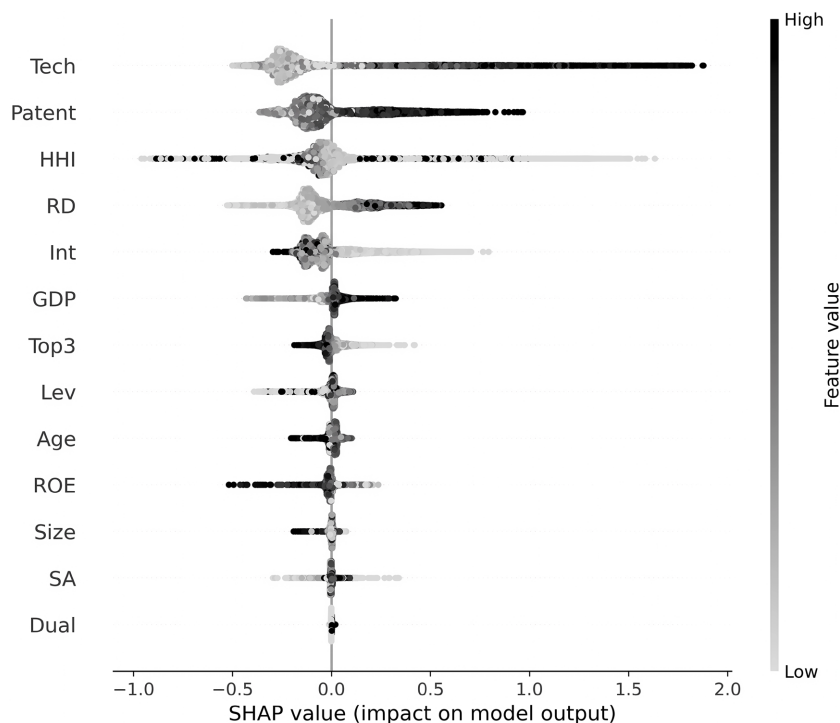


Figure 5. Summary plot of influencing factors (random forest)
图 5. 影响因素摘要图(随机森林)

4.4.3. 多次重抽样检验

为保证特征重要性排序的稳健性, 本研究采用 Repeated K-Fold 稳定性检验。与单次划分或普通重抽样相比, Repeated K-Fold 能在多组独立样本划分中反复训练模型, 从而有效降低随机划分带来的偏误。具体操作中, 以 5 折交叉验证重复 20 次, 共得到 100 次 XGBoost 训练结果, 并在每次训练中提取综合重要性得分排名前 3 的特征。随后统计各特征进入 Top3 的频率, 以评估其重要性的结构性稳定性。在每次抽样中, XGBoost 模型会重新训练并计算 Gain (分裂中的损失下降)、Weight (被选作分裂节点的次数)、Cover (分裂涉及的样本覆盖率)三类特征重要性指标, 通过加权组合形成综合得分。

根据重抽样稳定性检验结果可见(表 4), Tech 与 HHI 在 100 次训练中均稳定进入重要性 Top3, 且综合得分最高, 说明其对数字化转型具有最强且最稳健的解释力。Patent 与 RD 虽进入 Top3 的频率相对较低, 但综合得分仍处于较高水平, 属于次重要且在部分抽样中发挥显著作用的因素。整体来看, 模型识别出的关键影响因素具有良好的稳定性。

4.5. 研究结论与启示

4.5.1. 研究结论

企业数字化转型作为推动资本市场高质量发展的关键路径, 其影响因素及作用机制一直是学界与政策层面关注的核心议题。本文以企业数字化转型水平为研究对象, 采用多元线性回归为基准模型, 并进一步引入 LASSO、决策树、随机森林和 XGBoost 等多种机器学习方法, 对不同企业特征变量对数字化

Table 4. Re sampling inspection results
表 4. 重抽样检验结果

影响因素	Top3 次数	综合得分
Tech	100	14.02
HHI	100	13.60
Patent	63	7.54
RD	35	7.49
GDP	2	7.22

转型的影响进行系统分析。研究发现：① 模型预测性能比较：在多种机器学习算法中，XGBoost 在样本内拟合优度(0.8628)与样本外预测性能(0.5648)上均表现最优，显著优于传统回归方法，表明其更适用于捕捉企业数字化转型的复杂机制。② 核心驱动因素识别：根据 SHAP 值的重要性排序，技术人员占比(Tech)、行业竞争强度(HHI)和专利数量(Patent)是影响企业数字化转型最为关键的三个因素，其预测重要性显著高于其他变量。③ 技术人员占比的关键作用：技术人员占比对数字化转型的影响存在明显的“规模效应阈值”。当技术人员占比低于 30%时，其对数字化转型的推动作用有限甚至为负；一旦超过该阈值，其边际贡献显著增强，呈现出强劲的单调递增效应。④ 行业竞争强度的非线性影响：行业竞争强度(HHI)与数字化转型之间呈“倒 U 型”关系。在 HHI 低于 0.1 的充分竞争市场中，企业数字化转型动力不足；随着竞争适度集中(HHI 介于 0.1 至 0.25 之间)，数字化转型水平显著提升；而当 HHI 超过 0.25 后，促进作用趋于稳定。⑤ 专利数量的边际递减效应：专利数量对数字化转型始终具有正向影响，但存在边际效应递减特征。当专利数量低于 3.5 时，促进作用较弱；在 3.5 至 4.5 区间内推动效应显著增强；超过 4.5 后，增速放缓，表明单一专利积累并非数字化转型的持续动力。

4.5.2. 政策启示

结合上述研究结论，本文从政府监管、企业战略等方面提出以下政策启示，以期为推动企业数字化转型和资本市场高质量发展提供参考：

1. 完善政策支持体系，强化技术人力资本建设
- 研究结果显示，技术人员占比是影响企业数字化转型的最关键因素之一，政府应进一步完善数字经济相关政策，鼓励企业加大研发投入与技术人才引进力度，特别是通过税收减免、科研补贴、人才培养基金等政策工具，支持中小企业建立稳定的技术研发团队。同时，应推动产学研协同创新机制，构建高水平的数字化人才供给体系，提升企业整体技术吸收与创新能力。
2. 优化市场竞争环境，形成适度竞争格局
- 行业竞争强度对企业数字化转型呈现“倒 U 型”关系，表明过度分散或过度垄断的市场均不利于数字化进程。因此，监管部门应在反垄断与产业协同之间实现动态平衡，鼓励适度集中与有序竞争。可通过推动产业链整合、促进数字平台开放共享、加强公平竞争审查等措施，为企业创造健康的竞争环境，从而激发数字化转型的内生动力。
3. 鼓励创新成果转化，提升专利质量与应用效率
- 专利数量虽对数字化转型具有持续正向影响，但存在边际递减效应，说明仅依靠数量积累难以支撑长期转型。政策层面应从“数量导向”向“质量导向”转变，完善知识产权评价体系与激励机制，促进高价值专利的商业化与产业化应用。政府可支持企业建立开放式创新平台和数据共享机制，强化跨领域技术融合，推动创新成果在生产管理、供应链协同及智能决策中的实际落地。

4. 推动多维度数据治理与智能监管机制建设

机器学习模型显示,数字化转型具有复杂的非线性特征,传统监管方式难以精准捕捉企业数字化发展动态。建议监管机构运用人工智能、大数据分析等技术,建立动态监测与评估体系,对企业数字化水平、创新绩效及资本配置效率进行实时追踪,以形成更具前瞻性和针对性的政策响应机制。

参考文献

- [1] Ardolino, M., Rapaccini, M., Saccani, N., Gaiardelli, P., Crespi, G. and Ruggeri, C. (2017) The Role of Digital Technologies for the Service Transformation of Industrial Companies. *International Journal of Production Research*, **56**, 2116-2132. <https://doi.org/10.1080/00207543.2017.1324224>
- [2] 孟凡生, 赵刚. 传统制造向智能制造发展影响因素研究[J]. 科技进步与对策, 2018, 35(1): 66-72.
- [3] Christian, D., Matthias, H., Jochen, W., et al. (2017) How AUDI AG Established Big Data Analytics in Its Digital Transformation. *MIS Quarterly Executive*, **16**, 81-100.
- [4] 刘冀徽, 田青, 吴非. 董事长研发背景与企业数字化转型——来自中国上市企业年报文本大数据识别的经验证据[J]. 技术经济, 2022, 41(8): 60-69.
- [5] 谢康, 胡杨颂, 刘意, 等. 数据要素驱动企业高质量数字化转型——索菲亚智能制造纵向案例研究[J]. 管理评论, 2023, 35(2): 328-339.
- [6] 李兰, 董小英, 彭泗清, 等. 企业家在数字化转型中的战略选择与实践推进——2022·中国企业家成长与发展专题调查报告[J]. 南开管理评论, 2022, 25(5): 191-204.
- [7] 余典范, 王超, 陈磊. 政府补助、产业链协同与企业数字化[J]. 经济管理, 2022, 44(5): 63-82.
- [8] 邱洋冬. 网络基础设施建设驱动属地企业数字化转型——基于“宽带中国”试点政策的准自然实验[J]. 经济与管理, 2022, 36(4): 57-67.
- [9] 陈强. 机器学习及 Python 应用[M]. 北京: 高等教育出版社, 2021.
- [10] 李斌, 邵新月, 李玥阳. 机器学习驱动的基本面量化投资研究[J]. 中国工业经济, 2019(8): 61-79.
- [11] Bertomeu, J. (2020) Machine Learning Improves Accounting: Discussion, Implementation and Research Opportunities. *Review of Accounting Studies*, **25**, 1135-1155. <https://doi.org/10.1007/s11142-020-09554-9>
- [12] 吴非, 胡慧芷, 林慧妍, 等. 企业数字化转型与资本市场表现——来自股票流动性的经验证据[J]. 管理世界, 2021, 37(7): 130-144, 10.
- [13] 赵宸宇, 王文春, 李雪松. 数字化转型如何影响企业全要素生产率[J]. 财贸经济, 2021, 42(7): 114-129.