

分类号  
U D C

密级  
编号 10741

兰州财经大学  
LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 制造业数字化转型驱动因素研究  
——基于机器学习的证据

研究生姓名: 贾映彬

指导教师姓名、职称: 关辉国 教授

学科、专业名称: 工商管理 企业管理

研究方向: 公司治理

提交日期: 2024年5月30日

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 贾映彬 签字日期： 2024.5.30

导师签名： 关辉国 签字日期： 2024.5.30

## 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 贾映彬 签字日期： 2024.5.30

导师签名： 关辉国 签字日期： 2024.5.30

**Research on drivers of digital  
transformation in manufacturing  
enterprises -- Evidence based on machine  
learning**

**Candidate : Jia Ying Bin**

**Supervisor: Guan Hui Guo**

## 摘要

在数字化时代，制造业企业面临着转型为数字化企业的挑战，这成为提高企业竞争力的重要途径。然而，关于制造业数字化转型的驱动因素的研究尚不够充分，现有研究多是考虑单一因素对企业数字化转型的影响。本研究旨在深入了解影响制造企业数字化转型程度的关键影响因素及其重要性，通过综合内外部因素，构建一个全面的企业数字化转型程度驱动因素研究框架。

本文使用 2017 年至 2021 年上市制造业企业数据，建立了一个制造业数字化转型的预测模型，并运用多种机器学习算法，包括 Lasso 回归、支持向量机、随机森林和 LightGBM, XGBoost 进行建模分析。结果表明，随机森林和 LightGBM, XGBoost 在拟合效果上表现最佳。然后使用 LightGBM 算法模型，通过排列重要性和 SHAP 框架下的方法，对驱动因素进行特征重要性分析，定量评估了各因素对制造企业数字化转型程度的重要性。本文发现，数字技术投资、行业竞争强度、地区数字环境、高管风险偏好、知识密集度和管理者能力在众多模型中表现出对企业数字化转型有较高的贡献度。针对这些重要因素，利用 Shap value 量化了特征对预测值影响的大小和方向（正或负），并构建了可视化工具，包括摘要图和部份依赖图，并进一步分析了它们对企业数字化转型各个维度的具体影响及不同驱动因素间的交互效应。

本文的结果显示，在众多驱动因素中，数字技术投资是最重要的驱动因素。行业竞争强度和知识密集度均较高时，可能会削弱对数字化转型的正向贡献。研究还发现，高管风险偏好、地区数字环境和管理者能力也与数字化转型密切相关。本文的研究结果为制造业企业在数字化转型过程中提供了重要的决策参考，有助于企业更好地推进数字化转型，提高数字化水平，增强企业竞争力。企业应加大数字技术投资，关注行业竞争态势，合理配置知识密集度，优化高管结构，关注地区数字环境，提升管理者数字化能力，制定全面的数字化转型战略，以顺利推进数字化转型。本文的研究结论具有针对性，主要针对制造业，未来研究可扩展到其他行业，探讨不同行业数字化转型的差异性。

**关键词：**企业数字化转型 机器学习 制造业 上市公司

## Abstract

In the digital age, manufacturing enterprises are faced with the challenge of transforming into digital enterprises, which has become an important way to improve the competitiveness of enterprises. However, the research on the driving factors of the digital transformation of manufacturing industry is not enough, and most of the existing studies consider the impact of a single factor on the digital transformation of enterprises. The purpose of this study is to deeply understand the key factors affecting the degree of digital transformation of manufacturing enterprises and their importance, and to build a comprehensive research framework on the drivers of the degree of digital transformation of enterprises by integrating internal and external factors.

This paper uses the data of listed manufacturing enterprises from 2017 to 2021 to establish a prediction model for the digital transformation of manufacturing industry, and uses a variety of machine learning algorithms, including Lasso regression, support vector machine, random forest and LightGBM, XGBoost for modeling analysis. The results show that Random Forest, LightGBM and XGBoost have the best fitting effect. Then, the LightGBM algorithm model is used to conduct a feature importance analysis of the driving factors through the method of ranking importance and SHAP framework, and the importance of each factor to the degree of digital transformation of manufacturing enterprises is quantitatively assessed. This paper finds that digital technology investment, industry competition intensity, regional digital environment, executive risk appetite, knowledge intensity and manager ability have a high contribution to enterprise digital transformation in many models. In view of these important factors, Shap value is used to quantify the magnitude and direction (positive or negative) of the influence of features on the predicted

value, and visual tools are constructed, including summary chart and partial dependency chart, and their specific impact on various dimensions of enterprise digital transformation and the interaction effects among different drivers are further analyzed.

The results of this paper show that among the many drivers, investment in digital technologies is the most important driver. Higher levels of industry competition and knowledge intensity may undermine positive contributions to digital transformation. The study also found that executive risk appetite, regional digital environments, and manager competencies are also strongly related to digital transformation. The research results of this paper provide important decision-making reference for manufacturing enterprises in the process of digital transformation, and help enterprises to better promote digital transformation, improve digital level, and enhance enterprise competitiveness. Enterprises should increase investment in digital technology, pay attention to industry competition, rationally allocate knowledge intensity, optimize executive structure, pay attention to regional digital environment, improve managers' digital capabilities, and formulate comprehensive digital transformation strategies to smoothly promote digital transformation. The research conclusions of this paper are targeted, mainly for the manufacturing industry, and the future research can be extended to other industries to explore the differences of digital transformation.

**Keywords:** Digital transformation of enterprises; Machine learning; Manufacturing,; Publicly listed companies

# 目 录

<b>1 绪论</b>	<b>1</b>
1.1 研究背景	1
1.2 研究目的及意义	2
1.2.1 研究目的	2
1.2.2 研究意义	3
1.3 研究内容与研究框架	4
1.3.1 研究内容	4
1.3.2 研究方法	6
<b>2 相关文献综述</b>	<b>7</b>
2.1 关键术语界定	8
2.1.1 企业数字化转型的定义	8
2.1.2 企业数字化转型驱动因素研究	10
2.2 企业数字化转型驱动因素相关的理论基础	14
2.2.1 社会技术系统	14
2.2.2 资源依赖理论	15
2.2.3 TOE 框架	15
2.3 机器学习理论	16
2.3.1 机器学习相关综述	16
2.3.2 机器学习方法	18
2.3.3 特征重要性	19
<b>3 用于测度特征重要性的机器学习模型构建</b>	<b>20</b>
3.1 指标测量	22
3.2 样本选取和数据来源	28
3.3 数据采集和预处理	28
3.3.1 填补缺失值	28

3.3.2 缩尾和标准化.....	29
3.3.3 划分训练集和测试集.....	29
3.4 建立模型.....	30
3.4.1 模型选取.....	30
3.4.2 评价指标.....	30
3.4.3 企业数字化转型模型建立和评价.....	31
<b>4 基于机器学习模型的特征重要性分析 .....</b>	<b>33</b>
4.1 特征重要性分析方法选择.....	33
4.2 企业数字化转型驱动因素重要性分析.....	34
4.3 对重要因素影响的分析.....	38
4.3.1 摘要图.....	39
4.3.2 部分依赖图.....	40
<b>5 研究结论与展望 .....</b>	<b>48</b>
5.1 研究结论.....	48
5.2 管理启示.....	49
5.3 研究不足与展望.....	50
<b>参考文献.....</b>	<b>52</b>
<b>附录.....</b>	<b>59</b>
<b>后记.....</b>	<b>65</b>

# 1 绪论

## 1.1 研究背景

### （1）数字化转型的时代必然性

随着时代的发展信息技术和大数据的不断革新，信息技术与制造技术密不可分，虚拟空间不断发展，人工智能技术取得重大突破。为了推动信息技术在制造业中的应用以及在未来的发展中抓住机遇，英国、美国、德国、法国在内的多个发达国家，都在基于自身情况制定并执行了一系列战略计划。例如，英国在 2013 年提出了《英国工业 2050 战略》，指出信息技术在制造业的生产过程、技术人才培养和管理、产品供应链等方面具有重大意义，未来信息技术与产品乃至产品生产网络密不可分。美国在 2012 至 2013 年间推行了《美国先进制造业国家战略》和《美国制造业创新网络计划》，强调对先进制造业增加投资力度，建立专门研究机构，培养专业的技术性人才，为技术进步注入新的血液，在推动技术进步的基础上，推动制造业快速发展（陈醒，2019 年）。德国在 2011 年提出《德国工业 4.0》的基础上，于 2018 年进一步发布了《德国人工智能战略》，目的在于通过加强人工智能技术与制造业的融合促进制造业发展升级（李金华，2020 年）。

互联网技术和人工智能的快速发展，引发了经济结构和社会结构的深刻变革，这一变化是机遇也是挑战。为了抓住机遇迎接挑战，政府陆续出台了许多政策，如《智能制造发展规划（2016-2020）》和《工业互联网创新发展行动计划（2021-2023）》，在于促进数字经济快速发展，推动我国数字化转型，促进跨越式发展。截至 2022 年我国的数字经济规模已达到 50.2 万亿元，同比增长 10.3%，数字经济在 GDP 中的占比达到 41.5%。信息技术的进步不仅改变了社会的生产方式，而且使数字化转型成为企业发展的新趋势和新要求

### （2）制造业企业数字化转型的必然性

张培（2020）提出，从战略管理的层面来看，制造业要保持可持续发展就必须进行数字化转型。为了适应发展趋势，技术创新和产业结构调整是必由之路，为了生存和发展传统制造业必须进行数字化转型。本部分将分析传统制造业数字

化转型的必然性，从外部和内部两个维度进行探讨。

在内部环境方面：一是数字经济飞速发展的今天，在传统制造业的人口红利已经消失、资源效率低下、产业链不完整、年轻劳动力对制造业工作不感兴趣的前提下，数字化转型是企业降低成本，提高工作效率的有效途径。根据世界经济论坛的报告，企业 17.6% 的成本可以通过数字化转型来避免。二是传统制造业的核心竞争力正在从单纯的制造能力转向综合的制造、数字化、服务能力。企业需要实现技术开发的协同和动态化，生产方式的智能化、多样化和柔性化，合作模式的灵活化和网络化，并通过商业模式创新，把握未来市场的机遇。三是因为组织结构庞大、机构臃肿、人员复杂、管理观念落后、管理水平低下、生产模式单一、产品不具备多样性等原因，许多传统的制造业面临被淘汰的风险。然而，在数字化环境中。根据数据实现组织扁平化管理和工作协同，利用多样的数据生产令消费者需要的产品，同时一些简单重复的劳动可以交给人工智能技术完成，简化企业结构，提高工作效率，增强企业竞争力和市场适应性。

在外部环境方面：一是数字经济的迅猛发展。信息技术与众多行业的融合发展，推动新技术在制造业的产品研发、生产、销售、服务中普遍应用，在提高数据利用率和顾客参与度的基础上，促进了组织创新和企业互联互通，为制造业的发展迈出一大步（马化腾，2017）。二是市场需求的转变。随着数字技术的不断发展，市场需求更具有个性化和多样化，促使企业从产品导向转向客户导向，从注重对产品的研发生产到更注重客户需要什么样的产品，以满足市场需求。③政策的陆续发布。多个部门如工信部、发改委、国资委等，都发布了促进企业数字化转型的方案，旨在推进企业的数字化、网络化和智能化、信息化的发展。

## 1.2 研究目的及意义

### 1.2.1 研究目的

数字化转型已成为企业改革，发展新的生产力，促进自己商业模式变化，实现可持续发展的关键。本研究结合这一现实背景通过机器学习方法探讨驱动制造企业数字化转型的影响因素，从机器学习可解释性的角度出发，采用各类机器学

习模型对数据集建模并评估预测精度，以其他机器学习模型作为稳健性验证，对机器学习算法模型采用 SHAP 解释方法研究影响因素。

(1) 系统讨论了在制造企业数字化转型影响因素的可解释性，本文对于驱动因素间的重要性的研究，有助于缩短企业提升数字化水平的研究过程，拓展了企业数字化转型研究领域的研究思路。

(2) 相比于此前制造企业数字化转型影响因素研究中使用的逻辑回归和各类统计检验方法，本研究采用基于机器学习模型的 SHAP 方法探究驱动制造企业数字化转型核心影响因素，为后续的指标研究提供了新的可能。

(3) 在制造企业数字化转型影响因素的研究中，尚无研究者使用机器学习方法。企业数字化转型的过程是个黑盒，各项因素对驱动企业数字化的作用机理是复杂的，本文为了探究这一驱动机理，使用机器学习方法来学习各因素对企业数字化复杂的关系，其解释对于过往的研究模型更具有可靠的现实意义。

## 1.2.2 研究意义

### (1) 理论意义

本研究目的在于分析促进企业实现数字化的核心动因。尽管目前的研究文献已经对宏观政策、市场环境和公司内部因素如何作用于企业数字化进程及其效果进行了广泛的探讨，这些研究大多只集中于单一因素的考量，未能充分考虑不同因素间的相互影响。本研究提出，只有那些在众多因素综合作用下仍能显著影响数字化转型的因素，才能在实践中为企业转型提供关键的指导。

为了解决这一问题，本研究采用机器学习方法，在全面回顾现有文献的基础上，构建了一个包含多元因素的企业数字化转型驱动因素框架。具体步骤包括：首先，对现有文献进行了深入的回顾，总结了宏观政策、市场环境以及公司内部因素是如何影响数字化转型的。然后，本研究运用机器学习算法对数据进行深入分析，以揭示数字化转型与各种因素之间的关联性以及这些因素间的相互作用，进而对每个因素的影响程度进行量化评估。最后，通过机器学习的 SHAP 方法，利用摘要图和部分依赖图对关键因素与数字化转型之间的关系进行了可视化分析，以识别出促进数字化转型的关键因素。

本研究旨在回答：在众多影响企业数字化转型的因素中，哪些变量是政府和企业改善数字环境时应重点关注的。同时，本文还探讨了以往研究中较少涉及的问题，如如何在模型中纳入单因素与其他因素的交互作用，以及哪些因素对企业数字化转型水平具有显著影响。这些问题的解答对于深入理解企业数字化转型的驱动因素至关重要。

## （2）实践意义

对于政府机构来说，探究那些旨在推进企业数字化进程的财政措施是否始终发挥积极作用，其效果是否显著，以及在何种条件下能够最大化其促进作用，构成了一个持续的探究课题。本研究旨在通过数据分析，揭示企业的具体环境与支持政策如何影响制造业的数字化转型，并评估在多因素交织的背景下财政措施的有效性，从而为后续政策的制订与执行提供参考。找准驱动因素的重要因素，对政府机构极为关键，才能对政策实施有更进一步的调整与规划，才能更加发挥政府的调节作用实现有针对性的政策部署，并确保政策的正面影响。

从企业管理实践的角度来看，本文对关键驱动因素的研究有助于企业更快地确定提升数字化能力的路径。企业不仅能够根据自身状况识别出限制数字化进程的关键阶段和主要因素，而且能够依据这些因素的重要性来优先解决核心问题。例如，若企业的投资并未带来预期的效益或效益不佳，企业应从多个角度探究限制其数字化转型的因素，解决内部治理的不足，优化管理层结构，以及更有效地对员工负责等。

## 1.3 研究内容与研究框架

### 1.3.1 研究内容

第一章为绪论。论述了企业进行数字化转型的必然性，并且从微观和宏观的层面上分析了数字化转型的意义。在简述完研究背景之后，说明本文的研究目的及研究意义，对本论文的总框架有一个总体的概括。绘制了技术路线图，指出本文的创新点与不足。

第二章，针对企业数字化转型进行相关文献综述。首先先对企业数字化转型

进行一个明确的定义,在梳理了过往研究对于企业数字化转型的驱动因素的研究。在论述了企业数字化转型驱动因素相关的理论基础。然后针对目前机器学习方法在研究中的应用进行了梳理,并且简单介绍了一下机器学习方法和本文所使用的摘要图、部分依赖图以及算法模型等。

在第三章中,为了研究驱动企业数字化转型的驱动因素,探究哪些驱动因素是重要的,需要先构建一个测度特征重要性的机器学习模型。本文先简介了一下本文所选取的驱动因素指标,并且对其进行了测量。在数据采集完成之后,需要先对其进行预处理,首先要填补缺失值,然后对其进行缩尾处理和标准化处理。之后是建立模型,并且对不同的模型利用指标进行评价。最后选择表现较优的模型作为本文所使用的模型,去研究驱动因素的重要性。

第四章,在确定好使用的算法模型之后,开始分析驱动因素的重要性。首先对比那些算法模型下的特征重要性,如果某个因素在所有的模型中都显出重要的特征,这说明这个因素对企业数字化转型有较为重要的影响。在找到几个较为重要的驱动因素之后,对这些驱动因素进行分析,本文使用了摘要图和部分依赖图,以探究他们是如何影响企业数字化转型的。

第五章描述结论,根据本文的研究结果和找到的那些重要驱动因素,撰写研究结论以及管理启示,并对研究的不足进行展望。

以下为本文的技术路线图。

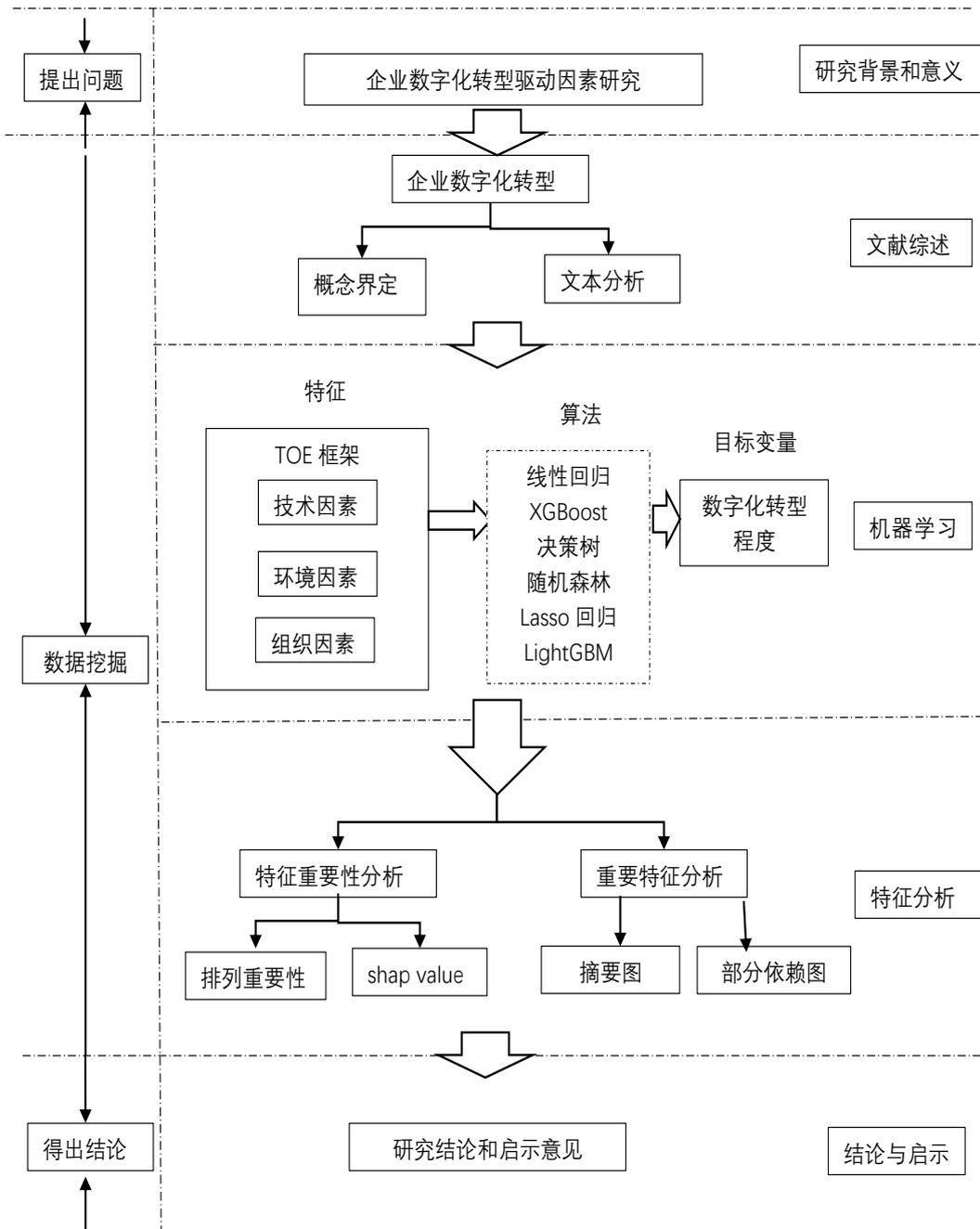


图 1.1 技术路线图

### 1.3.2 研究方法

#### (1) 定量研究法

本研究运用机器学习技术深入探讨了制造业数字化转型背后的多元因素。根据现有文献，推动企业数字化转型的各种因素可能与转型过程本身呈现出非线性

或复杂性的相互关系,并且这些因素间的相互作用显著影响着单一因素与数字化转型之间的关系。然而,传统的实证分析方法,主要用于探索自变量与因变量之间的线性关系,并不适合深入挖掘这种复杂的关系。尽管实证分析方法通过假设检验,可以较高概率地保证其结论的可靠性,但是对于多因素间的复杂规律,它并不适用。

本研究的目的是通过一个 TOE 框架,从技术组织环境三个层面上构建尽可能全面的驱动因素与企业数字化转型的模型。通过回归算法、支持向量机、决策树、随机森林、XGBoost 和 LightGBM 算法,用这些算法模型学习不同驱动因素间的复杂交互,揭示不同因素对企业数字化转型的影响。先对不同的算法进行评估,找出对预测企业数字化转型程度表现较好的几个算法。再用这些算法对驱动因素的重要性进行评估,然后利用部分依赖图和摘要图,展示重要驱动因素对企业数字化转型程度的影响方式和因素间的交互作用。

## (2) 文本分析法

在现行研究中,文本分析方法被广泛用于描绘企业数字化转型的进展。本研究采用文本挖掘技术和频率分析,深入分析了公开上市企业的年度报告,并以报告中数字化转型相关词汇的出现频率作为评估企业数字化水平的指标。具体而言,研究首先运用 Python 编程的网络爬取功能,整合了上海证券交易所和深圳证券交易所所有上市公司的年度报告,为后续的特征词汇筛选提供了数据支撑。随后,本研究回顾了众多聚焦于数字化转型的学术文献,提炼出相关的关键词汇。通过统计、匹配和计算这些关键词在年度报告中的出现次数,本研究对涉及关键技术领域的词汇进行了分类和汇总,最终构建了一个全面的评估企业数字化转型的指标体系。

## 2 相关文献综述

### 2.1 关键术语界定

#### 2.1.1 企业数字化转型的定义

##### (1) 数字化转型的定义

随着第四次工业革命的初步展开,企业和学术界对数字化转型的探索尚处于发展阶段,目前尚未形成普遍认可的定义。《传统产业数字化转型的模式和路径》研究报告由国务院发展研究中心发布,该报告强调数字化转型依赖于最新一代的信息技术,通过建立一套高效的数据收集、传输、处理和反馈机制,消解不同管理层级和行业间的信息障碍,从而全面提升产业运行效率并塑造新的数字经济发展格局。从具体来说,数字化转型通常包括了多个方面。首先是业务流程,数字化是通过信息技术对企业的内部和外部进行改造,使数字化贯穿企业的业务流程,使其更有效率。二是对组织结构进行调整,数字化转型使企业对目标发生了变化,而目标的变化则会促进组织结构的改变,使其更加适应数字化,更灵活的团队和更响应迅速的团队。三是数据出动决策,数字化转型带来了更多的数据和利用大数据分析的技术手段,从这些数据中可以提出与以往不同的洞察并改变企业的决策过程。四是商业模式也会发生改变,数字技术带来了新的商业模式,如平台化经营更加具有个性化的产品的。

我国也有许多学者对企业数字化转型进行了定义。如吴非等(2021)认为对企业进行数字化转型,本质上是对战略的变革,数字技术的改变与企业本身的业务资源组合配置,发挥了数据驱动效应,企业可以据此提升效率,优化结构。王子阳等(2020)认为数字化转型是指将数字技术在企业的经营与决策的全流程中进行运用。

在此基础上,本文认为企业数字化转型是对业务流程组织结构,决策过程以及商业模式的转型,目的是彻底改变生产过程和运营方式,从而有效提升生产效率和财务表现。

企业数字化转型具有如下特征:

一是数字化转型是一个持续的、演化的战略过程，要求企业不断进行策略更新、调整和优化。为了实现这一目标，企业必须制定出长期的转型规划。重要的是，企业领导者需要认识到数字化转型是一个逐步推进的过程，不应期望一开始就实现全面的转变。相反，应根据企业现有的人才以及财务情况，灵活的制定数字化转型策略。不能一味冒进，只追求单纯的转型，而忽视组织的目标去考虑。同时，要根据社会发展的趋势和业务需求的变化，不断地对转型战略进行相应的调整。

二是实现数据驱动运营模式是数字化转型的核心。在过往土地、劳动力、资本和企业家精神被称为关键的四大生产要素，但现在数据已经成为了关键的第五大生产要素。数据不仅影响了企业的决策过程，也是新兴的企业数据资产，并且可以对公司的资源进行有效的整理。挖掘，然后发现生产经营过程中出现的问题。并且数据是客观的，有直接性的，可以减少企业管理者决策的成本，为企业的经营提供有力的支持。

三是数字化转型是将数字技术与业务流程进行融合的过程。企业并非是为了转型而转型，而是源于新兴的市场需求以及不断发展的技术的双轮驱动。与信息化不同，数字化转型的特点在于业务与技术的深度整合。在此过程中，IT 部门不再仅仅是业务的支持者，而是需要与业务部门紧密合作，共同创造商业价值。这就要求企业内部的人才不能只具备单一的业务知识，更要掌握数字化的技术，或者组建跨领域的综合团队，实现业务与技术的高度融合。

## （2）关于制造业数字化转型的研究

国际学术界在探讨制造业数字化转型的过程中，关注技术在制造业转型过程中带来的影响。部分学者（Carmody P 和 Surborg B, 2014 等）认为，云技术以及网络技术能够改变过往的传统行业，推动企业业务的纵向发展。在产业转型面临困境的时候有积极作用。Wang Y (2007) 认为，信息技术及相关知识对企业数字化转型有重要的影响因素，传统的企业应该融入这类技术，以促进自身业务及结构调整。“互联网+”以及数字化技术推动传统产业的创新，针对原本的经营模式和业务流程以及生产方式进行了改造，推动了产业与新经济的快速发展。

针对互联网技术以及相关数字化技术的发展，大数据、云计算、人工智能等

新技术不断突破，传统制造业可以利用这些技术，改变自己的生产方式，提升生产效率。这与发达国家提出的“工业互联网”和“工业 4.0”概念不谋而合，全球各国都在引入新的技术和发展新的生产力去提升自己制造业的效率。已经完成工业化的国家通过发展信息技术，获得了完善而又庞大的技术产业，通过不同产业关联体系的基础建设使自己在产业处于全球领先的地位，以获取更多的利润。

(Dewan S, 2000)。以工业 4.0 为核心的新一轮工业革命是产业发展的重要机遇。国外提出了工业 4.0 的计划，这是针对于产业的新一轮工业革命，新的资源要素投入生产过程，新兴产业与互联网进行融合，实现了这些产业的共同发展。(Tang D, 2016)。互联网技术的蓬勃发展，促进了互联网相关应用在工业上的发展，工业与网络虚拟现实和网络物理系统整合在一起，使许多工厂实现了智能制造。

(Saldivar A A F, 2015; Jasperneite J, 2017; Gon P M, 2014)。

### 2.1.2 企业数字化转型驱动因素研究

技术-组织-环境 (TOE) 框架是一个全面的工具，众多学者已经采用 TOE 框架来探讨企业数字化转型的动因和内在机制。本文也将从技术因素、组织因素、环境因素这三个维度论述影响企业数字化转型的驱动因素。

#### (1) 技术因素

数字技术投资、数字技术管理能力以及能够快速适应环境的动态能力都是技术维度的重要因素。

在数字化转型中，企业的成功转型依赖于两大核心支柱：对数字技术的战略投资和管理能力的完善和创新。从战略层面看，如 VIAL G(2019) 等表明，传统制造企业在数字化转型中，显著体现在对数字化软硬件的投资深度上。此类投资不仅是资金的投入，更是对未来商业模式的深度思考。它们优化了企业的内部运作流程，大幅降低了信息搜寻、产品追踪及市场交易等环节的耗时与成本，赋予了企业更加敏捷的生产反应、精细的供应链协调与稳健的业务风险防线。因此，对数字技术的投资不仅是企业数字化转型的必然选择，更是其转型成功与否的衡量尺度。如李煜华(2022)所说的，企业为了满足软件和硬件方面的数字化需求，通过使用数字资源和数字技术，并购重组软件企业等方式，构建数字设施完善数

数字化支持。沈运红（2020）和綦良群（2021）都强调在企业数字化转型中，创新数字技术是关键，构建数字设施是支撑，推动企业数字化转型升级。

数字化转型是一场深刻的企业变革，它覆盖了从研发创新、产品设计、生产制造到市场营销的全价值链。在这场变革中，企业不仅要进行高额的数字技术投资，更要培养和积累与这些技术相匹配的管理智慧。这种技术管理能力，正是企业在识别、利用、整合乃至重塑数字技术方面的核心竞争力。在现实中，许多企业在数字化转型过程中遭遇了专业人才和管理经验的瓶颈，这使得数字技术难以真正融入企业的日常运营，从而影响了整体的经营效率。苏汝劫和常宇豪（2019）的研究指出，数字技术的综合运用能力是企业技术管理能力的核心，这种能力涵盖了数字技术的识别、利用、整合和重构。

此外，在数字经济时代，企业的动态能力已成为其智能升级的关键。Teece(1998)曾提出，动态能力是企业基于内外部资源整合，对外部环境变化做出快速适应的核心能力。随着研究的深入，Teece进一步将动态能力细化为感知市场变化的敏锐度、捕捉并获取有价值资源的能力，以及根据环境变化重构自身运营模式的灵活性。这些能力共同构成了企业在复杂多变的市场环境中持续发展的基石。那些拥有强大动态能力的企业，往往能够更好地感知市场风向，利用各类资源，从而在数字化转型的道路上稳健前行。

## （2）组织因素

在企业组织转型的过程中，有很多影响因素，其中，组织层面的因素尤为关键，它包括管理层的能力、知识密集度、高层管理者的风险倾向、资源基础以及企业在数字化转型上的关注分配等。

组织在数字化投入与产出差异的根源是人力资本，这一差异取决于管理层的技能和员工的知识水平（彭永涛，2022）。管理层的技能不仅仅是一系列的能力集合，更是一种内在的精神特质（邢文杰，2022）。这些能力，如风险决策、资源管理、机会识别等，在每一次战略抉择中都发挥着关键作用。它们相互作用，相互咬合，推动着企业不断前行。当这些能力得到充分发挥时，企业便能在数字化过程中保持平衡，实现高产出与低投入的完美融合（陈雪琴，2018）。在制造业的数字化转型中，高层管理者扮演着举足轻重的角色。他们不仅要为企业描绘

出清晰的数字化转型规划，还要敏锐地感知数字环境的变化，深刻理解数字技术的价值（孙磊，2020）。他们的能力，如数字化愿景的构建、业务创新与变革的推动、团队管理与组织发展的融合、组织进化的人才建设等，都是组织在数字化转型中不可或缺的资源（张路，2019）。

此外，数字化转型还要求企业组织结构实现从刚性“树状”向灵活“网状”的转变。在这种新型结构下，企业的每个节点都实现了横向连接和资源共享，企业的运营方式也从单点辐射、波浪式扩展转变为整体迭代。这种转变不仅打开了沟通渠道，让数字化转型成为全员参与的过程，也对组织的知识密集度提出了更高的要求。员工的知识技术水平成为了推动数字化转型的关键因素，组织知识密集度的提升将为企业在数字化道路上注入更强的动力（吴先明，2014）。因此，在数字化转型的道路上，组织需要不断提升自身的能力，培养具备数字化思维和技能的员工队伍，以应对日益复杂多变的市场环境。只有这样，企业才能在激烈的竞争中脱颖而出，实现持续稳健的发展。在 21 世纪的商业环境中，数字化转型已成为制造企业持续竞争力的核心要素。这种转变不仅涉及技术的升级，更涵盖了组织文化、管理理念和战略方向的全面革新。在这个过程中，管理者的能力和知识的密集度扮演着至关重要的角色。当企业内部知识存量丰富、质量高时，合作与创新的氛围自然更加浓厚。这种环境能够充分激发研发人员的创新思维，推动数字技术的持续进步。正如常嵘（2017）所指出的，知识密集度是推动企业数字化转型的重要动力。

而当转向组织层面时，管理者的角色不容忽视。他们不仅拥有决策权，更在塑造企业文化、推动战略执行方面发挥着关键作用。管理者的风险倾向差异，会在多个层面上影响数字化转型的路径。依据 Brickley 等（1988）的研究，可以将管理者分为两类：压力敏感型和压力抵抗型。压力敏感型管理者在面对挑战时往往表现得较为保守，这可能导致他们在执行外部治理职能时效果不佳。相反，压力抵抗型管理者则更加倾向于积极应对挑战，他们不仅致力于加强企业治理，还主动提高信息透明度，从而减少信息不对称问题，降低监督成本。这样的管理风格有助于企业将更多资源聚焦于数字化转型。在研发领域，压力抵抗型管理者展现出了更强的风险抵御能力和专业技能。他们不仅能够推动企业的研发工作，为

数字化创新营造有利环境，还能为转型提供必要的技术支持。这种全面的能力使得企业在数字化转型过程中更加稳健和高效（李华民，2021）。

通过明确管理者的角色和职责，以及合理规划和分配资源，企业可以更加稳健地迈向数字化未来。数字化转型是一项庞大的工程，它要求企业投入巨大的资源，并承担着高风险。这种转型的经济效益往往不是立竿见影的，这进一步增加了企业的融资成本和调整成本。因此，企业在推进数字化转型时，必须充分考虑自身的资源储备水平和资源运营能力。这些因素将直接决定企业能够走多远，实现多大的数字化转型（唐松，2020）。

数字化关注程度是另一个关键因素。它体现了企业对人工智能、大数据等数字技术的关注程度。这种关注度的分配在很大程度上影响着企业的行为选择。关注分配对企业战略决策的重要性不言而喻。当企业长期对数字技术保持高度的关注度时，它们的数字化转型实施程度通常会更高（李雷，2023）。

### （3）环境因素

环境维度的影响因素包括企业竞争压力、地区数字环境及环境不确定性。

在探讨制造企业数字化转型的影响因素时，不可忽视的是外部环境的多维度挑战。这些挑战来源于企业之间的竞争压力以及所处地区的数字化环境。在这个动态且充满竞争的商业生态系统中，企业与政府以及其他竞争实体之间的关系形成了一个错综复杂的网络（金珺，2020）。这种复杂的互动关系使得企业在资源获取和市场份额方面面临着巨大的竞争压力。这种竞争不仅源于企业内部资源的有限性，也受制于市场的局限性（李煜华，2022）。为了在这样的竞争环境中立足，领先的制造企业开始寻求数字化转型的解决方案。它们借助先进的信息技术和组织结构变革，以期在市场上保持持续的竞争优势。当竞争对手开始感受到这种数字化发展的压力时，它们也不得不跟随这种变革的步伐，以免被市场淘汰（金珺，2020）。

而地区的数字化环境对于制造企业的成功至关重要。数字基础设施、技术创新以及数字化的发展水平构成了地区数字化环境的核心要素（贾建锋，2022），它们不仅反映了当地的商业环境和经济水平，更为企业提供了转型所需的高端生产要素，如信息和数据。这些要素不仅助力企业从低端制造向高端制造的转型升

级，同时也为数字化创新产业集群的形成提供了有力支持，进一步推动了企业的转型升级进程（马中东，2020）。

当前我国各地区数字化环境的差异决定了不同地区的制造企业能否顺利实现数字化转型。这种差异不仅体现在数字化基础设施的完善程度，还体现在技术创新和数字化应用水平上。企业在制定数字化转型战略时，必须充分考虑所在地区的数字化环境，以确保转型的成功。随着数字化转型的深入，传统研究中关于组织变革的内部环境和外部环境界限变得模糊。数字化转型的影响因素不再局限于单一的环境维度，而是与内部环境、外部环境、行业特征以及消费者特征等多个因素相互交织、相互作用。这种复杂的关系要求企业在转型过程中必须具备全局视野和灵活应变能力，以应对来自各个维度的挑战和机遇。企业面临的竞争压力和所处的数字化环境是推动制造企业数字化转型的重要外部优势。

当前的环境不确定性正日益加剧，这无疑给企业的数字化转型带来了巨大挑战，但同时，这也成为了推动企业加速转型的强大动力（朱永明，2023）。另一方面，数字消费者的崛起和变化，正在推动企业进行数字化转型。随着消费者对数字技术的依赖日益加深，他们期待能随时随地访问虚拟资源，这也为企业的数字化转型提供了明确的方向和动力。在这样的背景下，企业不仅需要应对环境的不确定性，还需要满足数字消费者的需求，才能保持竞争力并实现可持续发展。

## 2.2 企业数字化转型驱动因素相关的理论基础

### 2.2.1 社会技术系统

组织理论中的一种视角是社会技术系统，它起源于英国人际关系研究所的E.L.Trist对达勒姆煤矿作业组织的研究。社会技术系统是一个由社会系统以及技术系统形成的一个组织。这个理论诠释了组织是包括了正式组织，非正式组织技术系统以及成员等多种不同的因素而构成的复杂系统。组织并非是由物质而组成，它包括了固定资产以及人力、机器、场地，也包括了人的行为。这个理论对构成了组织的许多复杂的因果关系进行了更为深入的探究，并更关注在组织中的人际关系系统，被称为现代的组织理论，与强调物质组织的传统组织理论相对应。数

数字化转型中各类因素有着广泛复杂的关系，有学者利用社会技术系统，从技术、任务、参与者和组织 4 个层面，梳理了哪些因素能够驱动数字化转型以及所带来的结果。（吴江，2021）。

### 2.2.2 资源依赖理论

资源依赖性理论占据了组织学研究领域的核心位置，资源依赖性理论是社会和组织理论中的一个概念，它描述了个体、群体或组织如何依赖于外部资源以维持其生存和发展。这一理论认为，资源是组织运作的基础，组织的目标实现和服务提供很大程度上取决于对各种资源的获取和利用。这些资源包括财务资源、人力资源、信息资源、物质资源等。在资源依赖性理论中，组织被看作是开放系统，它们必须与外部环境进行互动，以获取必要的资源。这种依赖性意味着组织需要建立和维护与其他组织的关系，这些关系可能包括合作伙伴关系、客户关系、供应关系等。

在企业向数字化转型迈进的过程中，资源的整合与战略的协同显得尤为重要。一方面，面对日益激烈的市场竞争，企业不论其数字化水平如何，都必须对创新技术、新兴业务领域以及关键员工的培训进行投资，以增强其市场竞争力。另一方面，数字化转型的过程通常伴随着较高的资本投入、潜在的投资风险以及收益的延迟实现，这些因素都会导致企业融资成本和调整成本的上升。因此，企业的资源储备能力及其资源管理效率成为影响数字化转型成功与否的关键因素。

### 2.2.3 TOE 框架

技术-组织-环境（TOE）框架是一个全面的工具，它基于技术、组织和环境三个核心维度，用于分析和解释企业采纳创新技术的动因。该框架的灵活性在于能够根据具体的研究需求和背景条件，调整不同的因素变量，使其成为适用于多种研究情境的“通用”理论。在制造业的数字化转型过程中，这三个层面的因素相互配合，共同作用于转型进程，使得 TOE 框架在分析此类转型时显示出其强大的解释力。

目前，众多学者已经采用 TOE 框架来探讨企业数字化转型的动因和内在机制。例如，赵艺婷（2022）基于 TOE 框架，从三个维度构建了一个用于分析制造企业数字化转型驱动因素的模型，并采用模糊集定性比较分析方法（fsQCA），从不同因素的组合角度出发，研究了这些因素如何共同推动制造企业的数字化转型。朱永明（2023）在 TOE 理论的基础上，运用大数据文本分析方法，描述了企业数字化转型的程度，并检验了资源基础、高管风险偏好等多因素对转型成效的影响，揭示了影响企业数字化转型的复杂因果机制。

综上所述，结合前述社会技术系统理论和资源依赖理论，本文将从技术-组织-环境三个维度选取指标。技术因素中包含了数字技术投资、数字技术管理能力、动态能力三个指标；组织因素中包含了注意力分配、资源基础、高管风险偏好、知识密集度、管理者能力五个指标；环境因素中包含了行业竞争强度、环境不确定性、地区数字环境三个指标。

## 2.3 机器学习理论

机器学习技术不仅在商业活动中应用广泛，也在学术领域中的应用逐渐日益增强。机器学习算法可以揭露数据间复杂的交互关系，现在新兴的算法可以提高模型的精准度，在预测分析领域也得到了广泛的应用。并且由于 SHAP 方法在机器学习中的广泛应用，对模型的可解释性也逐渐增强。

### 2.3.1 机器学习相关综述

#### （1）机器学习在企业研究领域的应用

在商业领域，机器学习技术已普遍渗透至诸如信贷评价、财务不端行为的侦查、资产价值评估、业务风险管理、利润调节以及资本配置等多项关键议题。丰富的学术研究围绕这些核心议题，通过发展多样化的模型以增强预测的准确性和决策的有效性。学者们经过对不同模型效能的细致对比，筛选出了性能最优的模型。例如，蒋盛益等人（2010）利用多种不同的算法，成功构建了一个针对财务风险的早期预警系统；方帅（2024）为了探究企业家在创新中起到的作用，用机

机器学习方法评估了董事长以及总经理在不同维度上的价值观的差异，并且利用这些差异对企业创新进行预测，识别出对企业创新最为重要的影响因素；吴世农（2023）构建了一个债券违约预警模型，基于财务信息和非财务信息使用机器学习方法；黄志刚等人（2020）对识别上市公司财务不端行为的前沿技术进行了全面的分析，并针对不同的算法构建出的模型进行了效果评估。

## （2）机器学习识别重要特征的文献

在跨学科研究领域，对于机器学习算法透明度的探索不断加深。对于预测性分析任务，一系列性能评估指标如准确率、混淆矩阵、ROC 曲线等已得到广泛应用。此外，在探讨因果关系和影响因素的学术文献中，逐渐采用了诸如特征显著性分析、依赖性图表（PDP）、个体条件期望图（ICE）以及 SHAP 等方法，以更深入地揭示模型的内在机制。多学科领域内，对机器学习算法透明度的追求正逐步增强。例如，孙发勤和冯锐（2019）基于神经网络与决策树技术，构建了一个模拟在线学习者成就的模型，并通过决策树的可视化分割过程来识别关键影响因素；张家普（2023）为了了解什么因素，对员工流失率的影响最为重要，用 XGBoost 模型预测员工流失，发现员工婚姻情况，所学习的专业领域对员工是否流失的影响比较重要；Parsa（2019）使用一个集成了交通、网络、人口统计、土地利用和气候等多个实时数据特征的模型来预测高速公路的事故发生率，并应用 SHAP 方法来解释这些特征的相对重要性；在探究影响病患医疗体验的各种因素及其相对影响力的研究中，佟金铎等（2017）采纳了基于机器学习的相关性分析方法；曹睿等（2021）利用 XGBoost 和随机森林模型对在线短租市场的定价影响因素进行了 SHAP 特征重要性分析，以识别关键因素，并为房东提供了提升收益的策略建议；张云中和秦艺源（2019）利用随机森林算法开发了一个标注系统标签模型，通过特征的信息增益来衡量其重要性，从而辨别出对产品质量具有显著作用的关键因素；朱振涛等（2020）结合 Lasso 正则化技术与随机森林的特征分析，研究了多个因素对消费者选择共享电动汽车意愿的影响。王志宁（2021）采用 XGBoost 算法预测员工流失率，并通过与支持向量机等多种模型比较，验证了 XGBoost 模型的优越性，同时使用 SHAP 方法分析了影响员工流失的主要因素。

近年来,随着企业数字化转型的加速推进,机器学习方法在该领域的研究展现出愈发显著的成果。多位学者从不同维度对此展开了深入研究。譬如,王新光(2022)亦运用文本分析与机器学习技术,深入剖析了管理者短视行为对企业数字化转型的潜在影响机制,并揭示出这种短视行为对数字化转型具有显著的制约作用。武常岐(2022)借助机器学习与文本分析技术,实证探讨了数字化转型进程中竞争战略选择对企业全要素生产率的深层次影响。此外,赵磊(2023)以制造业上市公司为研究样本,通过应用机器学习中的 KNN 算法,有效匹配了数字化转型程度指标,并深入研究了企业超常增长与数字化转型之间的复杂关系。研究结果表明,企业超常增长在一定程度上对数字化转型具有抑制效应,尤其是在订单式生产企业和离散型生产企业中,这种抑制效应表现得尤为突出。近年来关于数字化转型影响因素的研究日趋丰富,主要利用机器学习中的特征工程手段来评估各因素的重要性,并借助 SHAP 框架实现特征重要性的统一度量,还利用摘要图的方法确定对预测变量的影响程度。同时,部分学位论文还进一步运用了 PDP、SHAP 方法和 ICE 等工具,对模型进行了深入的解析与因素分析,从而更加全面地揭示了数字化转型的驱动机制。这些研究不仅为数字化转型的理论体系增添了新的内容,也为企业实施数字化转型提供了宝贵的参考依据。

### 2.3.2 机器学习方法

在回归问题上,机器学习算法主要用于预测或估计一个连续的数值型目标变量。这些算法可以从数据中学习特征与目标变量之间的关系,并利用这些关系来对新数据进行预测。常见的机器学习算法包括线性回归、决策树、随机森林、支持向量机、神经网络、梯度提升机等。这些算法在处理不同类型的数据和问题时各有优势,可以根据数据的特点和问题的需求选择合适的算法。在实际应用中,还可以通过调参和特征工程来优化模型的性能。调参可以帮助找到最佳的超参数设置,而特征工程则可以提取和选择对目标变量有更强预测能力的特征。本文选取了 Lasso 回归、随机森林、支持向量机、LightGBM 和 XGBoost 算法进行训练。

(1) Lasso 回归 (Lasso Regression): Lasso 回归是一种线性回归的变体,它通过在损失函数中加入 L1 正则化项(绝对值惩罚)来实现的。L1 正则化导致了

一些系数完全为零，从而实现了特征选择的效果。这使得 Lasso 回归特别适合于具有大量特征的模型，因为它可以自动进行特征选择，减少模型的复杂性。

(2) 支持向量机 (Support Vector Machine, SVM): 支持向量机是一种强大的监督学习算法，可以用于分类或回归任务。在回归任务中，它被称为支持向量回归 (SVR)。SVM 通过找到一个最佳的超平面来分隔数据，这个超平面可以最大化不同类别的边际距离。对于非线性问题，SVM 使用核技巧将输入空间映射到高维特征空间，使其能够处理非线性关系。

(3) 随机森林 (Random Forest): 随机森林由多个决策树组成。每个决策树都是基于随机样本和特征构建的。在回归任务中，随机森林通过平均所有决策树的预测结果来生成最终的预测。随机森林能够捕捉特征之间的复杂交互，并且对数据的分布不敏感，因此它们通常在回归问题中表现良好。

(4) LightGBM: 一个基于梯度提升框架的高效的机器学习算法。它是 XGBoost 的一个改进版本，旨在提高计算效率和减少内存使用。LightGBM 使用基于直方图的算法来加速训练过程，并且通过 leaf-wise 生长策略来提高模型的准确性。它在处理大规模数据集和具有大量特征的模型时表现优异。

(5) XGBoost (eXtreme Gradient Boosting): XGBoost 是一个流行的梯度提升框架，它通过优化损失函数并使用二阶导数来提高模型的性能。XGBoost 支持自定义损失函数，并且可以通过并行计算来加速训练过程。它在各种机器学习竞赛和实际应用中表现出色，特别是在处理结构化数据时。

### 2.3.3 特征重要性

特征重要性是本文研究的重点，度量不同变量对被预测变量的预测结果起到的作用。去探究哪些指标在影响企业数字化转型中起到重要作用。在训练模型时，不是所有的特征都同等重要，有些特征可能对预测结果有显著影响，而有些特征则可能作用不大。特征重要性可以帮助我们理解模型的工作原理，并在数据预处理、模型选择和解释性分析中发挥重要作用。本文将企业数字化转型程度视作被预测变量，其余指标被视为预测变量，寻找对模型预测效果影响较大的指标，分析哪些指标可以在多因素作用下依旧表现出重要的作用。并且对比分析不同算法

构建出的模型下的各指标重要性,如果某些指标在所有算法模型中都显示出共同的重要特征,则说明这些指标在算法模型中有重要作用,即这些指标对企业数字化转型程度有重要影响。本文将用两种方法分析特征重要性,并且利用 SHAP 方法中的摘要图和部分依赖图分析那些重要指标是如何影响企业数字化转型程度的。

### (1) 排列重要性

排列重要性 (Permutation Importance) 是一种后训练特征重要性评估方法,它通过评估特征值随机排列后模型性能的变化来衡量每个特征对模型预测的贡献。这种方法不依赖于模型的内部机制,因此可以应用于任何类型的模型,包括树模型、神经网络和支持向量机等。

排列重要性的计算步骤可以大概分为以下几步:①训练模型:使用全部特征训练一个机器学习模型。②性能基准:在验证集或测试集上评估模型的性能,得到一个性能基准。对于连续变量,可以使用均方误差(MSE)、均方根误差(RMSE)或平均绝对误差(MAE)等指标。③特征排列:对于每个特征,单独将其在验证集或测试集中的值随机排列。这样做会破坏该特征与目标变量之间的真实关系,而其他特征的值保持不变。④性能评估:使用排列后的特征集重新评估模型的性能。由于一个特征的排列不应该影响模型对其他特征的使用,如果模型的性能下降,这意味着被排列的特征对模型的预测能力是重要的。⑤重要性计算:计算模型性能下降的量,即排列后的性能与原始性能的差值。这个差值越大,说明该特征越重要。为了得到更稳定的重要性估计,可以多次重复排列和评估的过程,并计算平均性能下降。

### (2) SHAP 方法

SHAP (Shapley Additive Explanations) 方法是一种用于解释机器学习模型预测结果的有效工具。它基于博弈论中的 Shapley 值概念,能够量化每个特征对于模型输出的贡献程度。在解释一个模型的预测结果时,我们常常需要回答的问题是:每个输入特征对于预测结果的影响如何? SHAP 方法通过计算每个特征的 "Shapley 值" 来回答这个问题。Shapley 值的核心思想是,将特征视为参与一个 "博弈" 的玩家,每个玩家都会根据自己的特征值对最终的预测结果做出一定的 "贡献"

”。Shapley 值通过对所有可能的特征子集进行加权平均来计算每个特征的贡献值。SHAP 方法的应用非常广泛，特别是在解释复杂的黑盒模型（如深度神经网络）时很有用。通过使用 SHAP 方法，我们可以更好地理解模型的决策过程，找出哪些特征对于预测结果起到了决定性的作用，从而增加模型可解释性，并有助于发现数据中的一致性和不一致性。

具体来说，计算 Shapley 值的过程是通过迭代的方式进行的。首先，对于给定特征的可能取值，构建一个由其他特征组成的子集，并计算在这个子集中添加当前特征对于预测结果的贡献值。然后，通过对所有可能的子集进行加权平均，得到该特征的 Shapley 值。这个过程对于每个特征都会进行，从而得到特征对预测结果的贡献度量。SHAP 方法的一个重要特点是它能够处理任意类型的特征，包括数值型、分类型和文本型等。此外，SHAP 方法还提供了可视化工具，如摘要图和部分依赖图，以便更直观地理解特征的影响。

①摘要图（Summary Plot）：摘要图是 SHAP 方法提供的一种可视化工具，它将每个样本的 SHAP 值展示为一个点，点的位置由特征值决定，颜色表示特征值的大小。在摘要图中，水平轴表示特征的 SHAP 值，垂直轴表示特征本身。这样的布局可以看到每个特征是如何影响模型的预测的。水平轴上的点越远离原点，表示该特征对预测的影响越大。点的颜色表示特征值的大小，通常红色表示高特征值，蓝色表示低特征值。

②部分依赖图（Partial Dependence Plot, PDP）：部分依赖图是另一种可视化工具，用于展示一个或两个特征对模型预测的依赖关系。PDP 展示了特征的不同值如何影响模型的输出，但它不考虑其他特征的影响。在 PDP 中，可以在保持其他特征不变的情况下，观察一个特征的变化如何影响模型的预测。

本文将使用摘要图和部分依赖图，探究某些重要因素是如何影响企业数字化转型程度的。

### 3 用于测度特征重要性的机器学习模型构建

#### 3.1 指标测量

##### (1) 数字化转型程度

在探索企业如何转向数字化的过程中,这一转变被视为一项复杂的系统工程。尽管对于数字化程度的量化评估尚未形成统一规范,已有研究主要从宏观层面进行分析,或基于信息资源、数字化劳动力、信息系统应用等维度进行探讨。研究者们较常用可以直观表示的指标,比如无形资产比例,技术人员占比等。尽管这些指标具有一定的直观性,它们却未能全面展现企业数字化的整体状况。吴非等人(2021)提出了一种新的研究方法,运用文本挖掘技术来评估企业数字化转型的程度。他们的研究首先确定与企业数字化转型相关的关键词,并且使用爬虫来收集上市企业的年报,针对相关的关键词,提取文本数据,并排除包含“无”、“不”等否定意义的词汇以及与企业自身不相关的关键词(例如企业股东、客户、员工介绍等),最终将数字化相关的关键词频次作为衡量企业数字化转型程度的量化指标。

表 3.1 企业数字化转型的结构化特征词图谱

分类	维度	关键词
	人工智能技术	人工智能、商业智能、图像理解、投资决策辅助系统、智能数据分析、智能机器人、机器学习、深度学习、语义搜索、生物识别技术、人脸识别、语音识别、身份验证、自动驾驶、自然语言处理
底层技术运用	大数据技术	大数据、数据挖掘、文本挖掘、数据可视化、异构数据、征信、增强现实、混合现实、虚拟现实
	云计算技术	云计算、流计算、图计算、内存计算、多方安全计算、类脑计算、绿色计算、认知计算、融合架构、亿级并发、EB 级存储、物联网、信息物理系统
	区块链技术	区块链、数字货币、分布式计算、差分隐私技术、智能金融合约
数字技术运用		移动互联网、工业互联网、移动互联、互联网医疗、电子商务、移动支付、第三方支付、NFC 支付、智能能源、B2B、B2C、C2B、C2C、O2O、网联、智能穿戴、智慧农业、智能交通、智能医疗、智能客服、智能家居、智能投顾、智能文旅、智能环保、智能电网、智能营销、数字营销、无人零售、互联网金融、数字金融、Fintech、金融科技、量化金融、开放银行

### (2) 数字技术投资

本研究借鉴了傅国华（2023）的研究方法，将企业的数字化投资区分为硬件投资和软件投资两大类。其中，硬件投资主要涉及企业在电子设备、计算机办公设备等固定资产上的投入，而软件投资则主要涉及企业在信息系统和软件等无形资产上的投入。在收集到这两部分投资数据后，本研究将它们相加，并对相加后的总和进行自然对数变换，从而得到企业的数字化投资指标。

### (3) 数字技术管理能力

数字技术的综合运用能力是企业技术管理能力的核心,这种能力涵盖了数字技术的识别、利用、整合和重构。参考苏汝劼和常宇豪(2019)的研究,可以从以下三个维度来评估企业的技术管理能力:首先,技术创新的投资强度,可以通过研发支出占总资产的比例来衡量;其次,员工的知识水平,可以通过本科以上学历员工占总员工数的比例来衡量;最后,企业的资源管理和整合能力,可以通过净利润占总资产的比例来衡量。将这三个维度的指标进行标准化处理,并计算其平均值,可以得到企业的技术管理能力得分,得分越高,表示企业的技术管理能力越强。

#### (4) 资源基础

鉴于企业执行数字化转型所涉及的重大资本支出与资源密集度,仅有那些拥有充裕资源与充分能力的企业才能成功地进行这一转变。本研究采纳了陈庆江等研究者(2021年)的学术贡献,他们认为企业的现金流净值,作为经营活动的直接结果,提供了一个关于企业资源质量的清晰指标。基于此,本研究提出了一种新的度量方法,即通过比较企业在一个财务年度内经营活动产生的现金流量净额与其期末总资产的比率,以此来评估企业在该年度的资源基础。

#### (5) 注意力分配

企业的层面,数字技术的关注度往往不易直接量化。本研究采取组织职能的视角,通过观察企业是否创建与数字技术紧密联系的高级管理角色,来间接展示企业对数字化转型的关注水平。通过检查公开上市企业的年度报告,能够确认企业是否聘任了诸如首席技术官(CTO)、首席信息官(CIO)、首席数字官(CDO)等关键管理职位,这些职位对于推动数字化转型至关重要。如果企业任命了这些职位,则表明企业对数字化转型的关注程度较高,在分析中赋予其值为1;若未任命,则赋值为0。

#### (6) 高管风险偏好

在探究制造业企业高层管理人员对风险的态度时,本研究意识到这些态度可能对企业决策及最终的经营成效产生显著影响。具体而言,本研究借鉴了郭道燕(2016)的研究方法,从企业财务指标的五个维度,从资产结构维度选取了风险资产占总资产的比重,从偿债能力维度选取了资产负债率,从盈利结构维度选取

了核心盈利比率，从利润分配维度选取了留存收益率，在现金流量维度选取了自身资金满足率以及资本支出率，其中盈利结构维度和利润分配维度为逆向指标，进行了负号处理。具体指标见表 3.2。

表 3.2 高管风险偏好评价指标

评价内容	指标名称	符号	解释说明
资产结构	风险资产占总资产的比重	X1	(交易性金融资产+应收账款+可供出售金融资产+持有至到期投资+投资性房地产)/总资产
偿债能力	资产负债率	X2	总负债/总资产
盈利结构	核心盈利比率	X3	-主营业务收入/(主营业务收入+营业外收入+公允价值变动损益+投资收益)
利润分配	留存收益率	X4	-(一般盈余公积+任意盈余公积+公益金+未分配利润)/净利润
现金流量	自身资金满足率	X5	-(经营活动现金流入+期初现金及现金等价物)/所有现金流出
	资本支出率	X6	构建固定资产、无形资产和其他长期资产所支付的现金/总资产

### (7) 知识密集度

企业技术知识的运用通常是通过资本投入或劳动力技能来实现的，这在高技能人才方面尤为明显。因此，本研究通过分析劳动力中技术知识的构成，即企业技术人员占员工总数的比例，来量化企业的技术知识集中度。

### (8) 行业竞争强度

本研究采纳了钟熙等人（2021 年）提出的方法，使用赫芬达尔-赫希曼指数（HHI）来评估市场竞争的强度。HHI 指数的较低，意味着市场竞争更为激烈。为了使分析结果便于计算，本文中样本企业所在行业的竞争强度，用 1 减去 HHI 的数值表示。

### (9) 管理者能力

本研究采纳了 Dermerjian 等人（2012 年）的研究方法，结合 DEA 和 Tobit 模型来解析企业管理者对企业整体效率的具体影响。研究首先基于 DEA 模型，通过计算各企业的综合效率来评估其运营表现。在此过程中，将固定资产净额 (PPE)、无形资产净额(Intan)、商誉(Goodwill)、研发支出(R&D)、营业成本(COGS)、销售与管理费用(SG&A)作为输入变量，同时将营业收入(Sales)作为输出变量，以此来估算企业的效率得分。

$$Score = \frac{Sales}{axPPE + xIntan + xGoodwill + xR\&D + xCOGS + xSGA}$$

接下来，本研究进一步利用 Tobit 模型来找出管理者对企业效率的具体贡献。模型中控制了企业层面的多个影响因素，包括企业规模、市场份额、自由现金流、成立时长、国际化水平以及业务多元化水平。通过回归分析后，模型的残差被用来代表管理者的能力水平。

#### （10）环境不确定性

由于环境不确定性的特性和数据库中对其的直接衡量指标的缺失，本研究采纳了 Ghosh（2009）提出的方法来评估样本企业所经历的环境不确定性水平。首先本研究使用 STATA 软件对企业的营业收入与当年的时间进行 OLS 回归分析。回归的残差即为异常销售收入的波动。波动值与销售收入的平均之比可以估算，样本企业当年未经调整的环境不确定性。然后未经行业调整的环境不确定性水平的中位数，作为该行业环境不确定性的基准值。然后，通过比较未经行业调整的环境不确定性与行业环境不确定性基准值，得出经行业调整的环境不确定性水平。应用上述步骤，本研究利用 2017 至 2021 年的样本数据，计算出每年样本企业的环境不确定性水平，并最终与各年度的平均值作为样本企业当年环境不确定性的结果。

#### （11）地区数字环境

在遵循王军等人（2021）的研究路径基础上，本研究综合了 2017 至 2021 年期间中国 31 个省份的横向数据，构建了一个评价区域数字发展水平的指标框架，并采用熵权法对其各项指标进行权重分配。这一评价体系涵盖了互联网接入节点数量、移动通信基站规模、专利授权状况等多个层面的因素。通过这些指标，可以综合评估各地区在数字经济发展方面的相对优势和未来潜力，从而为政府政策

制定和企业投资提供重要参考依据。对于部分数据的缺失，本研究运用插值法进行了有效的填补处理。

### （12）其他指标

为保证指标的选取尽可能全面，还需要一些研究驱动企业数字化转型因素的常用指标。本文参照赵宸宇（2021）的研究，添加其控制变量作为本文的指标。包括企业规模、企业年龄、资产负债率、股权集中度、流动比例、总资产收益率、所有权性质。

所有指标及其描述如下：

表 3.3 各维度指标与指标描述

项目	指标	指标描述
数字化转型	数字化转型程度	使用文本分析测度
技术因素	技术管理能力	3 个维度指标标准化后取均值
	数字技术投资	无形资产中软硬件投资
	注意力分配	企业是否设置与数字技术相关的高管职位
	资源基础	现金流量净额与期末总资产之比
组织因素	高管风险偏好	主成分分析方法提取反映高管风险偏好水平的主成分
	知识密集度	企业技术人员与员工总数之比
	管理者能力	从企业全效率中分离出管理者的贡献度
	行业竞争强度	1 减去 HHI 指数后的差值
环境因素	环境不确定性	企业经行业调整后年度均值
	地区数字环境	熵值法拟合当地数字基建指标
	企业规模	企业总资产的对数值
	企业年龄	当年年份 - 成立年份 + 1
治理因素	资产负债率	总负债/总资产
	股权集中度	前 5 位大股东持股比例之和
	流动比例	流动资产/流动负债
	总资产收益率	净利润/资产总额
	所有权性质	国有企业记为 1，非国有企业记为 0

## 3.2 样本选取和数据来源

本文以制造业上市企业作为研究对象，制造业在全球经济中占据重要地位，其数字化转型不仅影响企业本身，还关系到整个产业链的升级和国家的经济结构调整。因此，研究制造企业的数字化转型可以较好地反映和理解企业数字化转型的普遍规律。推动制造业高质量发展是建设我国现代化经济体系的内在要求。通过分析制造企业的转型实践，可以揭示数字化转型的内在动因和成功因素。

中共十九大报告提出,加快建设制造强国,加快发展先进制造业,推动互联网、大数据、人工智能和实体经济深度融合,在中高端消费、创新引领、绿色低碳、共享经济、现代供应链、人力资本服务等领域培育新增长点、形成新动能。习近平总书记也在党的二十大报告中指出:“坚持把发展经济的着力点放在实体经济上”,“促进数字经济和实体经济深度融合”。2017年,中国企业开始广泛实施数字化转型。因此,本文研究时段为2017—2021年。企业进行数字化转型的成果并非是能立刻看见的,需要一定的时间,因此本文驱动因素指标选取的时间段,是预测指标即企业数字化转型程度的前一年。即自变量为2017—2021年数据,结果变量数字化转型程度为2018—2022年数据。本文样本企业为制造业上市公司,剔除缺失值过多样本后,本文研究样本一共有12374行。本文中企业数字化转型程度指标通过文本分析企业年报所得,各项指标及各指标下属维度的指标数据均来自国泰安数据库(CSMAR)。

## 3.3 数据采集和预处理

### 3.3.1 填补缺失值

因为本文所选取的驱动因素较多,在原始数据中存在大量的缺失值。为保证研究的质量和准确性,本文对缺失率较高的数据行直接删除,对缺失率不高的数据行利用不同方法进行数据的填补。

(1) 首先横向观察数据行,如果某样本企业在当年度的数据缺失很多,比率高于30%即删除此数据行。

(2)对于数据缺失不太多的企业,则先观察它的指标在前两年是否有数据,如果有数据,则利用 Python 对数据进行企填补,即利用前后两年的数值线性填充中间年份的数据;

(3)分析数据缺失情况,发现仍有大部分数据未填充,于是采用机器学习方法中的 KNN 预测值填充,即利用相似公司的该指标值进行填充;

(4)在经历完上述的不足后,大部分指标已经完成填补。不过上述处理方式都是针对驱动因素指标的填补,而对于衡量企业数字化水平的指标,如果缺失则直接删除此数据行。

### 3.3.2 缩尾和标准化

经过以上数据处理后,2017年至2021年的数据共有12374行。为保证数据有效,需要消除一些极端值对于结果的影响,进行了上下1%的缩尾处理。同时因为有些数据的绝对值过大,为保证所有数据的数值在同一维度下,对数据进行了标准化处理,均值为0,标准差为1。

### 3.3.3 划分训练集和测试集

在机器学习中,划分训练集和测试集是一个重要的步骤。训练集是用来训练机器学习模型的数据集。通过从训练集中学习,模型能够学习到数据中的模式和规律,从而建立一个能够对未知数据进行预测的模型。训练集通常包含了大量的样本数据,这些数据涵盖了模型可能遇到的各种情况。测试集是用来评估训练后模型性能的数据集。测试集应当与训练集独立,即在模型训练过程中未使用过。通过测试集,我们可以客观地评价模型的预测能力,了解模型在未知数据上的表现。本文将数据按照8:2的比例划分为训练集合测试集。

## 3.4 建立模型

### 3.4.1 模型选取

为保证建立的模型足够可靠,本文选取了不同类型的机器学习模型用于测量各个特征在预测企业数字化转型程度上的重要性。机器学习之前需要先将各指标统一标准,本文采用 Z-Score 方法进行标准化。为了保证模型的选取,尽可能的全面。本文主要选取了两种模型,一个是基于数据进行距离计算的算法包括普通线性回归、Lasso 回归、支持向量机;另一个是基于树的模型包括 LightGBM、XGBoost。本文利用这 6 个算法模型,逐个建立企业数字化转型程度的评估模型,然后对这 6 个模型进行评价。

### 3.4.2 评价指标

在数字化转型模型的建立中,我们需要选择合适的评价指标来衡量模型的性能。以下是一些常用的评价指标,它们提供了对模型预测准确性和拟合程度的深刻洞察。

#### (1) 均方根误差 (RMSE)

均方根误差是模型预测值与实际观测值之间差异的平方的平均值的平方根。RMSE 衡量了模型的预测误差的标准差,越小表示模型预测越准确。

$$\text{公式: } RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

其中,  $n$  是样本数量,  $y_i$  是实际观测值,  $\hat{y}_i$  是模型的预测值。

#### (2) 决定系数 (R-squared)

决定系数是一个用于衡量模型拟合程度的指标,取值范围在 0 到 1 之间。值越接近 1,表示模型能够更好地解释目标变量的变异性。

$$\text{公式: } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

#### (3) 平均绝对误差 (MAE)

平均绝对误差是模型预测值与实际观测值之间差异的绝对值的平均值。

MAE 度量了模型的平均预测误差，越小表示模型越准确。

$$\text{公式: } MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

这些评价指标提供了对模型性能的全面评估，帮助确定模型的适用性以及数字化转型中驱动因素的建模过程中的效果。在模型选择和调整中，这些指标将成为改进模型性能的重要工具。

### 3.4.3 企业数字化转型模型建立和评价

本文的代码的运行都在 Anaconda jupyter 中进行，并调用了 Scikit-learn、pandas、LightGBM、xgboost、SHAP 等机器学习库。

基于制造业企业 2017 年至 2021 年共计 12374 行样本数据建立普通线性回归、Lasso 回归、支持向量机、随机森林、LightGBM、XGBoost 模型。模型的参数也会对模型的表现有重要的影响，为保证能够使用最优参数，因此在建模过程中使用网格搜索调参优化，采用最优参数进行后续分析，各模型的参数如下：

表 3.4 数字化转型程度模型参数

模型	参数
线性回归	fit_intercept=False
Lasso 回归	alpha= 0.01
支持向量机	C= 1, kernel= 'rbf'
随机森林	max_depth= None, n_estimators= 200
LightGBM	learning_rate=0.1, num_leaves= 300
XGBoost	learning_rate= 0.1, max_depth= 9

本文使用 RMSE、MAE、R2 这三个评价指标，分析这 6 个算法模型，在训练集和测试集上的表现，结果如下表所示：

表 3.5 数字化转型程度模型效果

模型	R2_train	R2_test	RMSE_train	RMSE_test	MAE_train	MAE_test
线性回归	0.2007	0.1577	0.8964	0.9077	0.7427	0.751
Lasso 回归	0.1979	0.1609	0.898	0.906	0.7457	0.7516
支持向量机	0.4361	0.2598	0.753	0.8509	0.5763	0.6841
随机森林	0.915	0.3629	0.2923	0.7895	0.2373	0.6464
LightGBM	0.9808	0.3919	0.1388	0.7712	0.108	0.6242
XGBoost	0.907	0.3558	0.3057	0.7939	0.2285	0.6447

从上表中可以看出随机森林、LightGBM、XGBoost 在模型的表现上显著优于线性回归、Lasso 回归、支持向量机。随机森林、LightGBM、XGBoost 这三种基于树的模型在训练集的拟合度都达到了 0.9 以上,在测试集也达到了 0.35 以上。从 RMSE、R2、MAE 这三个指标在不同算法模型的测试集和训练集综合来看,LightGBM 是表现最好的算法,因此本文在后面使用摘要图和部分依赖图进行分析的时候,会使用这个算法来构建模型。

## 4 基于机器学习模型的特征重要性分析

上一章节对不同算法建立的模型进行了评估。本章则基于上一章表现较为优异的算法模型来评估企业数字化转型程度的特征重要性。本文主要利用两个方法来进行重要性的评估，一个是排列重要性，另一个则是 SHAP 方法中的  $mean(|shap\ value|)$ ，综合这两个指标找出影响企业数字化转型程度最为重要的因素。

### 4.1 特征重要性分析方法选择

特征重要性分析的原理可以这样理解，在一个预测模型中，不同的特征对最终预测结果的贡献是不一样的。有的特征可能非常重要，而有的特征可能根本不重要或者重要性很小。因此，识别哪些特征对预测结果影响最大，对于提高模型的性能和理解数据的内在结构非常重要。特征重要性分析方法的选择取决于数据类型、模型算法、业务需求和解释的深度。例如，对于需要理解特征之间复杂交互关系的复杂模型，SHAP 是一个很好的选择；而对于需要快速、简单解释的模型，决策树或随机森林提供的特征重要性评估可能就足够了。本文的被预测变量是企业数字化转型程度，如果某一个驱动因素对预测结果的影响最大，则可以视为这个因素，它对于企业数字化转型影响程度是同样重要的。

特征重要性评估的方法主要分为三类：过滤法（Filter Methods）、包裹法（Wrapper Methods）和嵌入法（Embedded Methods）。过滤法是一种相对较简单的特征选择方法。它不依赖于模型，而是在特征子集的选择之前单独评估每个特征的重要性。这类方法通常会计算特征的相关性、信息增益或基于模型的评分来评估特征的重要性，并根据这些评分来选择特征。过滤法的主要优点是它不依赖于特定的模型，因此选择的特征对于任何模型都是通用的。但是，它可能不会考虑到特征之间可能存在的交互作用。包裹法是一种更进阶的特征选择方法，它依赖于特定的模型。在这种方法中，特征子集的选择是在模型训练的过程中完成的。通常，所有可能的特征子集都会被考虑，并且每个子集都会被用来训练一个模型，然后根据模型的性能来评估特征子集的质量。包裹法可以选择出对特定模型最有

利的特征子集，但是它的计算成本通常较高，尤其是当特征数量较多时。嵌入法是一种将特征选择过程与模型训练过程相结合的方法。在这种方法中，特征的重要性是在模型训练的过程中动态评估的。模型在训练时会记录下每个特征对预测的贡献度量，这些度量可以直接用来评估特征的重要性。嵌入法的优点是它不需要单独的特征选择步骤，但是它对模型的依赖性较强，选择的特征可能只对特定的模型有效。

同时由于本文的目标是探索全部特征的重要性，嵌入法在模型训练过程中评估特征重要性，可以直接在模型中计算每个特征的贡献。对于全特征重要性评估，可以在模型训练后使用某些算法（如 SHAP）来分析每个特征对预测的贡献。在嵌入法中，具有全局一致性的是排列重要性和 SHAP 方法，因此本文使用这两个方法去研究企业数字化转型驱动因素的重要性。

## 4.2 企业数字化转型驱动因素重要性分析

本文接下来将对制造企业数字化转型程度进行分析，上文中那 6 个算法模型中表现最好的三个算法是随机森林、LightGBM、XGBoost。本文将使用这三个算法构建模型，评估方法包括排列重要性和  $mean(|shap\ value|)$ ，排列重要性用 `sklearn.inspection` 的 `permutation_importance` 函数。后者调用 `Shap` 包计算 `Shap value` 后进行绝对值和平均值计算得到。以 XGBoost 模型的排列重要性分析方法为例，其代码和运行结果如所示：

```
# 特征重要性
feature_importance = xgb_model.feature_importances_

# 打印特征重要性
print("\nXGBoost Feature Importance:")
for feature, importance in zip(X.columns, feature_importance):
    print(f"{feature}: {importance}")
```

XGBoost Feature Importance:  
数字技术投资: 0.0638476386662979  
数字技术管理能力: 0.03368552774190903  
注意力分配: 0.02767694741487503  
资源基础: 0.03142417594790459  
高管风险偏好: 0.045901209115982056  
知识密集度: 0.15874479711055756  
行业竞争强度: 0.1027849093079567  
管理者能力: 0.050314512103796005  
环境不确定性: 0.034707821905612946  
地区数字环境: 0.06366987526416779  
企业规模: 0.05204694718122482  
企业年龄: 0.04464751482009888  
资产负债率: 0.03805612027645111  
股权集中度: 0.05079865828156471  
流动比例: 0.04807185009121895  
总资产收益率: 0.03835088014602661  
所有权性质: 0.11527062952518463

图 4.1 XGBoost 模型-Permutation Importance 方法得到的重要因素

本文将找出不同算法，不同评估方法下重要性排名前 10 的驱动因素。即对制造企业数字化转型程度最重要的 10 个影响因素，看看究竟是哪些驱动因素，在不同算法不同评估方法下，均展现出了较为重要的特性。

表 4.1 不同模型下的企业数字化转型程度影响因素重要性

排名	随机森林-per	随机森林-shap	LGBM-per	LGBM-shap	XGBoost-per	XGBoost-shap
1	知识密集度	知识密集度	数字技术投资	行业竞争强度	知识密集度	行业竞争强度
2	数字技术投资	数字技术投资	数字技术管理能力	知识密集度	所有权性质	知识密集度
3	行业竞争强度	行业竞争强度	注意力分配	数字技术投资	行业竞争强度	数字技术投资
4	地区数字环境	地区数字环境	资源基础	高管风险偏好	数字技术投资	高管风险偏好
5	股权集中度	高管风险偏好	高管风险偏好	地区数字环境	地区数字环境	地区数字环境
6	企业规模	管理者能力	知识密集度	企业规模	企业规模	企业规模
7	管理者能力	所有权性质	行业竞争强度	所有权性质	股权集中度	数字技术管理能力
8	高管风险偏好	企业规模	管理者能力	管理者能力	管理者能力	管理者能力
9	所有权性质	股权集中度	环境不确定性	流动比例	流动比例	所有权性质
10	数字技术管理能力	流动比例	地区数字环境	股权集中度	高管风险偏好	流动比例

为了全面揭示特征的重要性，本文采用了两种不同的特征重要性评估方法，并将多个模型输出的结果进行了综合，计算出全局贡献度。通过对这些结果进行聚合分析，并采用了一种类似于投票机制的方法，以统计和确定排名前 10 的关键因素。如表：

表 4.2 投票得出的影响企业数字化转型程度的影响因素

因素指标	全局贡献度	投票结果	因素指标	全局贡献度	投票结果
数字技术投资	13.08%	6	流动比例	4.71%	4
行业竞争强度	8.98%	6	数字技术管理能力	4.61%	3
地区数字环境	8.37%	6	企业年龄	3.89%	0
高管风险偏好	8.33%	6	总资产收益率	3.76%	0
知识密集度	7.60%	6	环境不确定性	3.74%	1
所有权性质	6.94%	5	资产负债率	3.62%	0
企业规模	6.18%	5	注意力分配	2.76%	1
管理者能力	6.17%	6	资源基础	2.05%	1
股权集中度	5.21%	4			

从上表中可以看出，前 10 个特征中有 6 个是全部认同的，在不同算法模型以及不同的重要性评估方法中。其中数字技术投资在所有特征最为重要，在所有特征对企业数字化转型程度的影响中贡献度高达 13.08%，即企业数字化转型程度最受数字技术投资的影响，即数字技术投资的水平最能影响企业数字化转型程度，同时数字技术投资的大小会对企业数字化转型程度产生大的波动，可能的原因是数字技术作为现代信息技术的前沿，具有高风险、高收益的特点。企业在数字化转型过程中，需要大量投入资金进行技术研发、设备升级、人才引进等，这些投入不仅需要企业具备足够的资金实力，还需要对市场趋势、技术发展有敏锐的洞察力和判断力。如果企业不能准确把握市场和技术发展的方向，或者投入不足，可能会导致数字化转型的失败，从而对企业的发展产生重大影响；其次是行业竞争强度，高度竞争的市场环境中，企业需要不断提升自身的竞争力，通过数字化转型可以更好地满足客户需求，提高生产效率，降低成本等，从而在竞争中获得优势。如果行业竞争激烈，企业更需要加快数字化转型的步伐，提升自身的竞争力；然后，地区数字环境也是企业数字化转型程度的一个重要影响因素，良好的地区数字环境可以为企业提供更好的数字化基础设施、技术支持和人才储备，从而促进企业数字化转型的进程。例如，发达地区的互联网、云计算、大数据等

技术更为成熟，可以为企业提供更高效、更稳定的技术服务，帮助企业更好地实现数字化转型。高管风险偏好也是影响企业数字化转型程度的一个重要因素。高管的个人风险偏好会对企业的战略决策和投资行为产生影响，从而影响数字化转型的进程。如果高管倾向于风险规避，可能会对数字化转型的风险和不确定性持谨慎态度，从而减少对数字化转型的投入和推进力度。相反，如果高管具有较高的风险承受能力，可能会更加积极地推进数字化转型，以寻求更高的收益和增长机会。

知识密集度指的是企业研究人员的占比。企业的知识密集度较高，即拥有较多的研究人员时，企业可以更快地获取和吸收新的知识，包括数字技术、市场趋势等。这些知识能够为企业提供更多的创新机会，帮助企业更好地应对市场变化，推动数字化转型的进程。其次，管理者能力也是影响数字化转型的关键因素之一。数字化转型是一项复杂的系统工程，需要管理者具备战略规划、组织协调、资源整合、变革管理和创新思维等多方面的能力。如果管理者能力不足，可能会导致数字化转型的推进缓慢或出现偏差。因此，企业需要加强管理者的培训和选拔，提升管理者的能力和素质，以支撑数字化转型的顺利进行。

值得注意的是，所有权性质影响企业数字化转型程度也较为显著。国有企业由于其特殊的所有权性质，通常拥有更多的资源和资金，可以投入更多的资源进行数字化转型，从而更快地实现数字化转型。其次，国有企业通常承担着更多的社会责任和政策任务，因此更有可能得到政府支持和政策优惠，这对于数字化转型也是非常有利的。然而，所有权性质也可能会对企业的数字化转型产生一定的限制。例如，国有企业的决策流程可能更加复杂和繁琐，导致数字化转型的推进速度较慢。此外，国有企业的管理层可能更加稳定，但也可能导致数字化转型的战略和方向不够灵活和应变能力不足。下文的部分依赖图可以更好的解释和观察所有权对企业数字化转型程度的影响。

### 4.3 对重要因素影响的分析

上一章中已经找到了对企业数字化转型程度较为重要的6个因素。但是目前并不知道这些因素是如何影响企业数字化转型的，并不知道它们的影响是正或负

大或小，以及不同因素间是否有交互作用。SHAP 结合了几种现有方法，创建了一种直观、理论上合理的方法来解释任何模型的预测。Shap value 量化了特征对预测值影响的大小和方向（正或负），框架基于 Shap value 构建了可视化工具，包括摘要图、部份依赖图、力图等。摘要图通常用于展示模型的一些关键统计信息，如特征的分布、特征间的相关性等。它可以帮助我们快速了解数据的整体结构和特征的重要程度。在某些情况下，摘要图也可以用来比较不同模型之间的特征重要性。

### 4.3.1 摘要图

SHAP 方法的摘要图是一种用于解释机器学习模型预测结果的图形工具。它结合了特征重要度和特征的影响，通过颜色、位置和重叠程度等视觉元素，展示了每个特征对预测结果的贡献程度。摘要图中，每个点代表一个特征和一个实例的 Shapley 值。y 轴上的位置由特征决定，x 轴上的位置由 Shapley 值决定。颜色从低到高代表特征的值，重叠点在 y 轴方向上是有抖动的，因此可以了解每个特征的 Shapley 值的分布情况。通过观察摘要图，可以了解特征对预测结果的影响程度，并确定哪些特征对模型的预测结果具有显著影响。摘要图还提供了对模型行为的全面理解，而不仅仅是单一的解释。它可以帮助理解模型中的特征如何相互关联，以及它们如何共同影响预测结果。

本文的摘要图基于 lightGBM 模型，如图 所示。从摘要图可以看出，行业竞争强度、所有权性质和总资产收益率较为显著的呈现为偏负向关联指标，也就是取值越高该企业数字化转型程度也就越低。而知识密集度、数字技术投资、高管风险偏好、地区数字环境和企业规模这五个指标较为显著的呈现为偏正向关联指标，也就是取值越高该企业数字化转型程度也就越高。

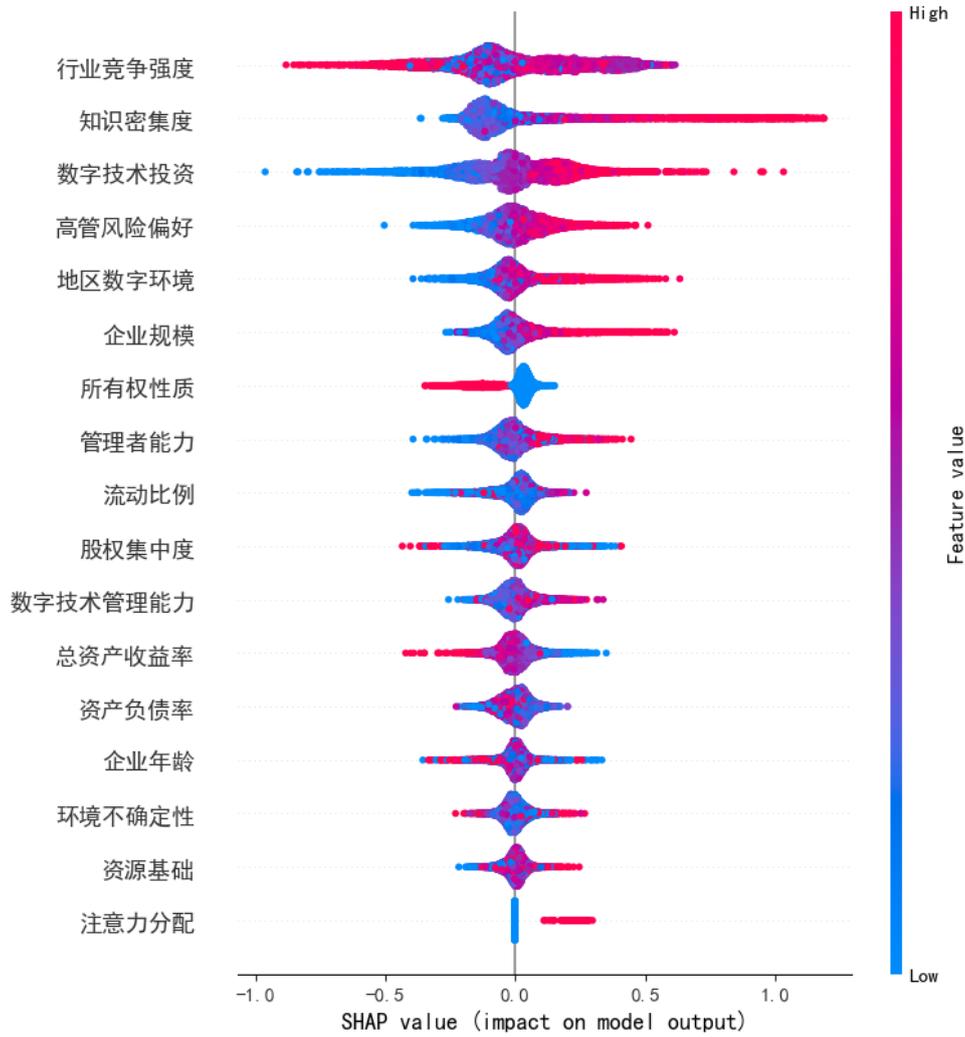


图 4.2 摘要图

### 4.3.2 部分依赖图

部份依赖图展示一个特征或两个最具交互效应的特征对模型预测结果的边际效应，可以显示目标和特征之间的关系是线性的、单调的还是更复杂的。部分依赖图将特征对目标的边际效应用散点图的形式呈现，本文用其解释各重要特征对企业数字化转型程度的影响。在上文中发现数字技术投资、行业竞争强度、地区数字环境、高管风险偏好、知识密集度这六个指标在多个模型中都表现出较为重要的特性，现将利用部分依赖图对这些指标进行分析。

#### (1) 数字技术投资

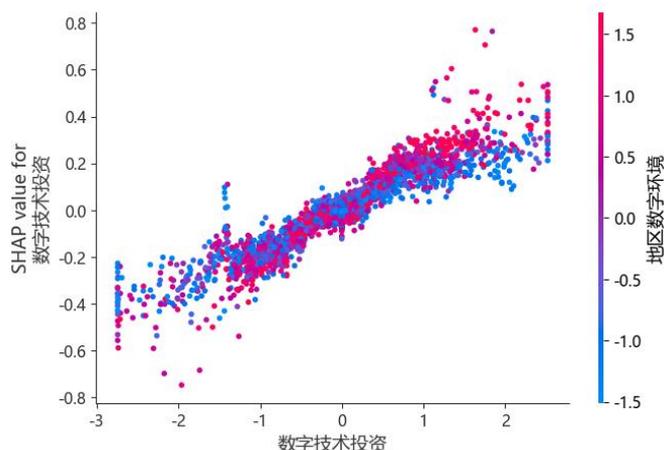


图 4.3 数字技术投资-地区数字环境

数字技术投资是指企业在信息技术和数字解决方案上的投资，包括硬件、软件、服务、人才等方面的投入。这些投资旨在提高企业的运营效率、增强客户体验、优化业务流程、创新产品和服务，以及开拓新的业务模式。企业数字技术投资可能包括但不限于云计算、大数据、人工智能、物联网、区块链等前沿技术的应用。数字技术投资是企业在进行数字化转型的过程中最直观的反应。企业数字技术投资是企业实现数字化转型的关键驱动力，它为企业的数字化转型提供了必要的资源和支持。

从这一部分依赖图可以看出与数字技术投资交互效应最大的指标是地区数字环境，并且随着数字技术投资的逐渐增大，它对企业数字化转型程度的贡献也就越大。地区数字环境涵盖了多个方面，包括但不限于基础设施的数字化程度、政府支持政策、数字技术人才的供给以及数字化市场的成熟度等。这些因素综合影响了企业在数字化转型过程中的投资决策和实施效果。地区的数字化基础设施水平对企业数字技术投资具有直接的影响。高度发达的数字基础设施通常意味着更为便捷、高效的数字化操作，使得企业能够更顺利地采用先进的数字技术，从而提高生产力和竞争力。地区数字环境中数字技术人才的供给情况也对企业的数字化投资产生深远的影响。充足的数字技术人才储备能够确保企业在数字化转型中能够顺利吸纳并运用最新的技术成果，推动创新和技术演进。因此，企业在选择数字化转型投资时，往往会考虑地区内数字技术人才的培养和引进情况。数字

化市场的成熟度也是企业数字技术投资的关键因素之一。成熟的数字化市场往往伴随着更广泛的数字化应用场景和更丰富的数字化服务供给，企业在这样的市场中更容易找到适应自身需求的数字解决方案，也更有可能是通过数字化手段拓展业务、提升客户体验。

## (2) 行业竞争强度

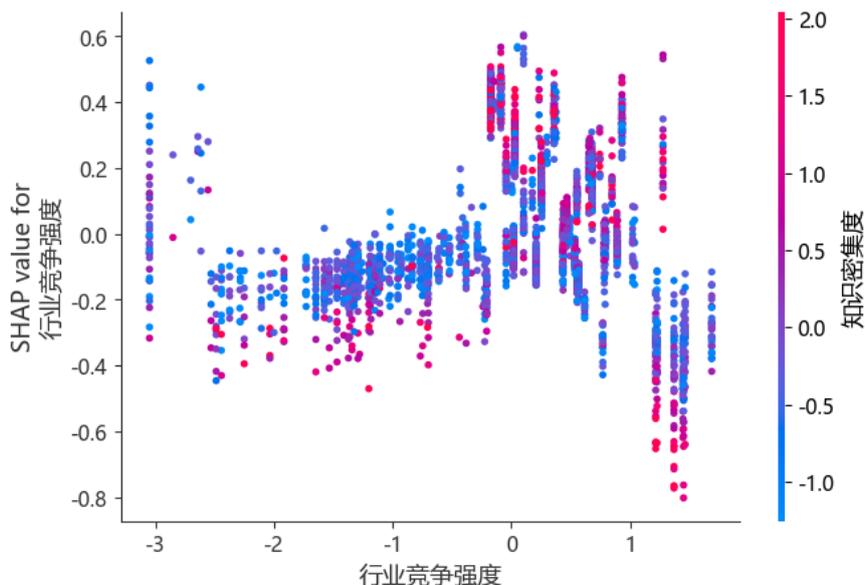


图 4.4 行业竞争强度-知识密集度

行业竞争强度是指特定行业内企业之间竞争的激烈程度，它反映了行业内企业在争夺市场份额、客户、资源和利润等方面的竞争状况。行业竞争强度受多种因素影响，包括行业内企业的数量、市场增长率、产品差异化程度、行业进入和退出壁垒、消费者对价格的敏感度、供应链结构、技术变革速度等。

通过对企业所在行业竞争强度与数字化转型程度的关系进行检验，本文进一步探讨了行业竞争强度对企业数字化转型的影响。企业所在行业竞争强度对企业数字化转型程度的影响如图所示。可以看到与行业竞争强度交互效应最大的指标是知识密集度，这一关联性反映了在数字化转型的过程中，知识密集度在高竞争行业中可能发挥着重要作用。在高竞争强度的行业中，企业面临更大的压力，需要通过数字化转型来提高效率和灵活性，以更好地应对竞争对手的策略和市场变化。

当企业所在行业竞争强度较低时，观察到企业的数字化转型程度相对较低。这可能是因为在竞争相对较弱的行业环境中，企业可能感受到较少的数字化压力，缺乏足够的动力去推动数字化转型。此时，企业可能更倾向于保持传统的经营模式，而非采取更为先进的数字技术。然而，当行业竞争强度增加时，我们观察到对企业数字化转型程度有一定的正向影响。这表明在较为竞争激烈的行业环境中，数字化转型可能成为企业获取竞争优势的一种手段。通过采用数字技术，企业可以提高生产效率、优化业务流程，并更好地适应市场的变化，从而在激烈的竞争中脱颖而出。值得注意的是，当行业竞争强度过高时，对企业数字化转型程度的增强作用可能会减弱。这可能是因为在极度竞争的环境下，企业可能陷入短期竞争和成本压力，导致其在数字化转型上投入不足。因此，在行业竞争强度极高的情况下，企业可能更注重应对当下的竞争挑战，而相对忽视了长期的数字化发展战略。

观察图像的右下角，有个值得注意的地方，即当行业竞争强度和知识密集度均较高时，观察到对企业数字化转型的贡献率为负并达到最大，这一现象提供了有趣的见解。这可能反映了在极高的行业竞争和知识密集度的双重压力下，数字化转型可能面临一些独特的挑战和限制。当行业竞争强度和知识密集度都很高时，企业可能面临更为激烈的市场竞争和更加复杂的技术环境。这可能导致企业在数字化转型过程中面临更高的成本、更复杂的技术整合问题以及更加困难的创新挑战。此外，极高的行业竞争和知识密集度可能使得企业在数字化转型过程中难以找到差异化和创新的空间，因为许多竞争对手可能也在追求相似的数字化战略。这种同质化竞争可能削弱了数字化转型的实际价值，导致负向的效应。

### （3）地区数字环境

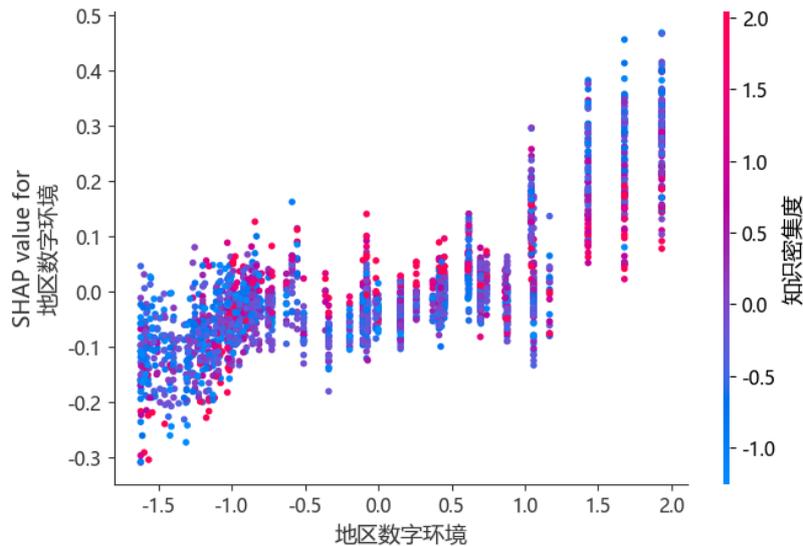


图 4.5 地区数字环境-知识密集度

地区数字环境是指一个特定地区内数字技术和信息通信技术的基础设施的综合体现。它包括但不限于互联网普及率、移动网络覆盖等。良好的数字基础设施是企业进行数字化转型的前提。地区数字环境中的网络、云计算、数据中心等基础设施的完善程度直接影响到企业数字化转型的基础能力和成本。

地区数字环境则表现出与 shap 值整体正相关的趋势，即地区数字环境越高，其企业数字化转型程度也就越高。并且与地区数字环境交互效应最大的指标，同样也是知识密集度。首先，地区数字环境对企业数字化转型程度的正相关趋势表明，地区的数字化基础设施以及其他数字化环境因素，对企业采用和应用数字技术起到积极推动作用。高水平的地区数字环境为企业提供了更为便捷和支持的数字化条件，使得企业更愿意进行数字化投资并更容易实施数字化转型战略。

其次，知识密集度再次显现为一个重要的交互因子，与地区数字环境之间存在显著的关系。这表明在数字化转型的过程中，知识密集度不仅在行业竞争中起到关键作用，同时在地区数字环境中也发挥着关键的连接作用。高知识密集度可能意味着更多的数字技术专业人才和创新资源，这进一步促使企业更积极地参与数字化转型，并更好地利用地区数字环境所提供的机会。

#### (4) 高管风险偏好

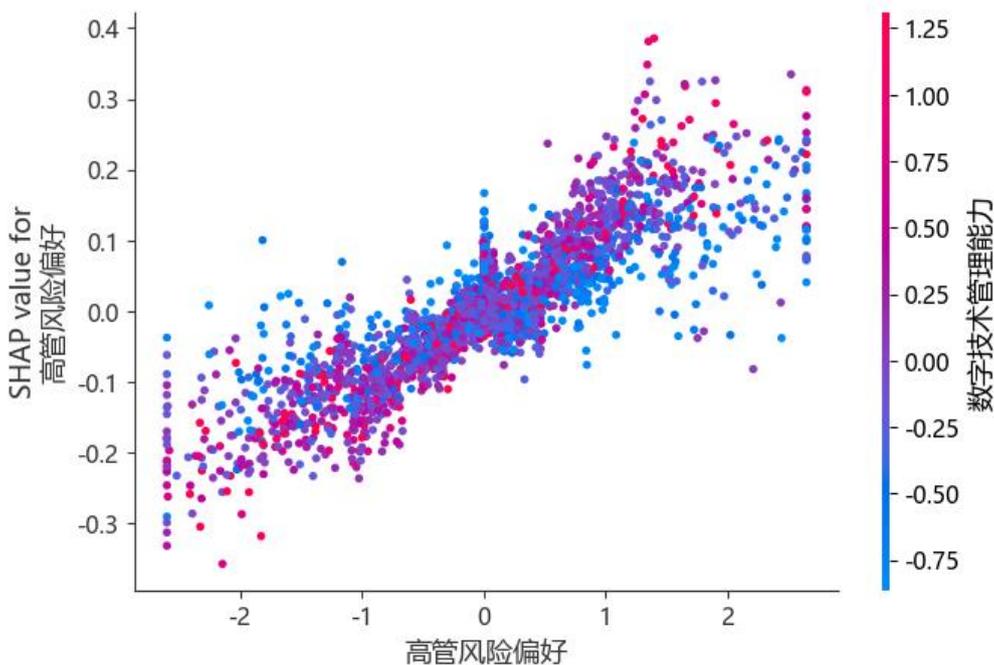


图 4.6 高管风险偏好-数字技术管理能力

在图中观察到高管风险偏好与企业数字化转型之间呈现正向线性关系的趋势，这一发现强调了高管对于风险的积极态度在数字化转型中的重要作用。该趋势的线性关系表明，随着高管风险偏好的增加，企业对于数字化转型的投入和效果呈现逐渐增强的态势。高管风险偏好的增加可能表明企业领导层对于外部环境变化更为敏感，更愿意承担数字化转型所带来的潜在风险。这种积极的风险态度有助于企业更加果断地采取数字化转型策略，尤其是在面对不确定性和变革时。

值得注意的是，高管风险偏好与数字技术管理能力之间存在最大的交互效应，这也在一定程度上反映了高管对于数字技术管理能力的重视。数字技术管理能力是指高管对于组织内数字技术的规划、整合和运营的能力。高管在数字化转型中的积极参与和领导，以及对数字技术管理的高效能力，对于数字化转型的成功具有关键性的影响。

在数字化时代，高管的数字技术管理能力不仅关系到企业数字化战略的执行效果，还直接影响着组织的创新能力和未来竞争力。因此，数字技术管理能力与高管风险偏好之间的交互效应的增强，表明在数字化转型中，高管需要更加注重数字技术领域的管理实践，以更好地引导企业应对数字化带来的挑战和机遇。

#### (5) 知识密集度

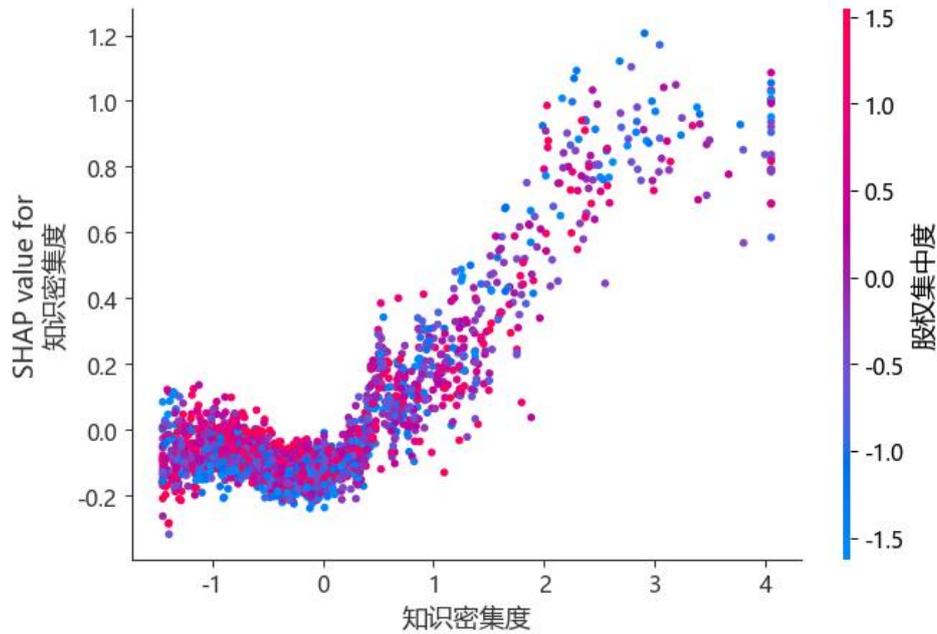


图 4.7 知识密集度-股权集中度

观察到知识密集度对于数字化转型程度的正向相关趋势，以及与知识密集度最强的交互效应是股权集中度。首先，知识密集度的正向相关趋势表明，随着知识密集度的增加，企业数字化转型程度呈现逐渐增高的态势。这可能反映了知识密集的行业或组织更容易吸引和培养高水平的数字技术人才，从而推动数字化转型的深入实施。高水平的知识密集度可能为企业提供了更多的创新和技术驱动力，有助于加速数字化转型的步伐。其次，与知识密集度最强的交互效应是股权集中度。这一发现指示了在数字化转型过程中，股权集中度可能对知识密集度与数字化转型之间的关系起到调节和加强的作用。股权集中度高的企业更容易实施一揽子的数字化战略，因为高度集中的所有权结构可能有助于更迅速地制定和执行数字化战略。高股权集中度可能意味着决策的集中和执行的高效，有助于企业更加果断地应对数字化挑战。在数字化转型的复杂环境中，需要迅速的决策和执行能力，而高股权集中度可能提供了这样的组织结构条件。这一交互效应的强化可能还意味着在高知识密集度的环境中，股权集中度的提高更为显著地推动了数字化转型。这可能是因为在知识密集的领域，数字技术的引入和应用更需要有力的领导和决策支持，而高股权集中度有助于确保这种支持的有效传导。

## (6) 管理者能力

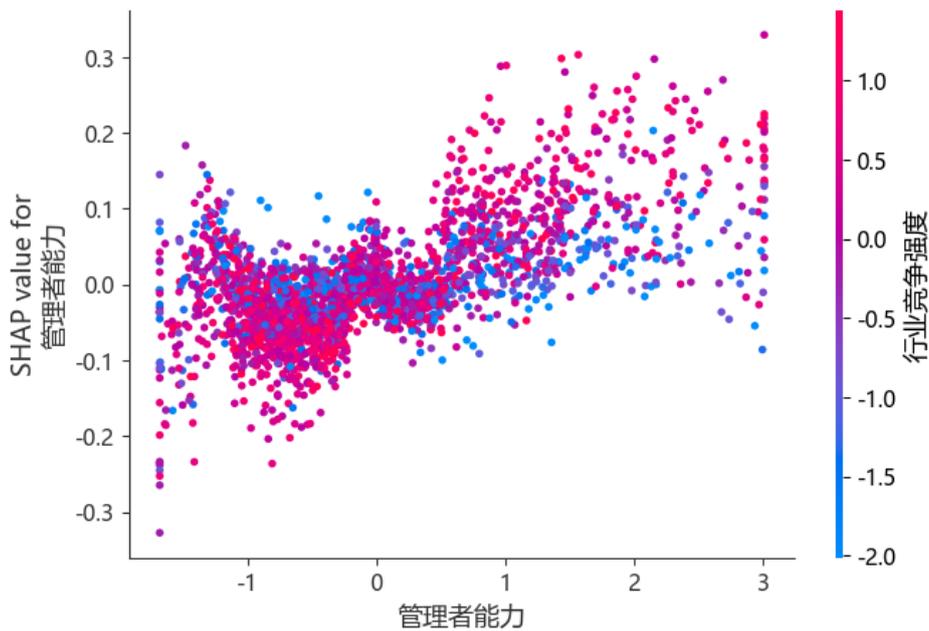


图 4.8 管理者能力-行业竞争强度

由图可知，与管理者能力交互效应最大的是行业竞争强度，以及在不同管理者能力水平和行业竞争强度条件下对数字化转型贡献率的值得关注的现象。

首先，当管理者能力较弱的情况下，高行业竞争强度似乎削弱了对企业数字化转型的贡献率。这可能反映了在竞争激烈的行业环境中，对于数字化转型的成功需要更为高效的领导和决策能力。管理者在应对强烈竞争时，需要更好地制定和执行数字化战略，以迅速适应市场变化。因此，当管理者能力较弱时，行业竞争强度可能对数字化转型的贡献产生负面影响。

相反，当管理者能力较强的情况下，高行业竞争强度似乎增强了对企业数字化转型的贡献率。这可能反映了在具备较强领导力和决策能力的管理者指导下，企业更能够应对激烈竞争环境中的数字化挑战。优秀的管理者能够更好地引导组织应对市场变化，制定切实可行的数字化战略，并有效执行这些战略以取得竞争优势。这一现象还可能与管理者能力在数字化转型中的引导作用有关。在强烈竞争的环境中，管理者能力的提升可能推动企业更加积极地采用先进的数字技术，更灵活地调整战略以适应变化的市场需求。

## 5 研究结论与展望

### 5.1 研究结论

本文研究发现。随机森林、LightGBM 和 XGBoost 这三个算法模型在预测数字化转型程度方面有较好的表现。本文使用排列重要性和 SHAP 方法去评估不同驱动因素的重要，结果表示数字技术投资、行业竞争强度、地区数字环境、高管风险偏好、知识密集度以及管理者能力在企业数字化转型的过程中比较重要。而其中数字技术投资是最重要的影响因素。以下部分针对得到的结论进行以下分析：

(1) 在对数字化转型的影响因素进行模型比较的过程中，结果显示基于树的模型，如随机森林、LightGBM 和 XGBoost，在预测数字化转型程度方面取得了较好的效果。随机森林、LightGBM 和 XGBoost 都属于基于树的集成学习模型，这些模型以其在处理非线性关系、高维特征和复杂交互关系方面的强大能力而闻名。在数字化转型这一复杂多变的场景中，通常存在大量的非线性关系和特征之间的复杂交互，因此基于树的模型能够更好地捕捉到这些复杂性，从而提高预测的准确性。其次，这些基于树的模型在处理大规模数据集时表现优异。在数字化转型研究中，通常会涉及到大量的数据，包括来自不同领域的多种指标和因素。基于树的模型能够有效地处理这些复杂数据结构，同时具备较高的计算效率，使其成为适用于数字化转型研究的理想选择。

(2) 数字技术投资对企业数字化转型的影响至关重要，是影响企业数字化转型程度最重要的因素。这主要是因为数字技术投资直接反映了企业在数字化转型中的投入力度，是推动企业数字化转型最直观、最直接的方式之一。数字技术投资可以提升企业的生产效率和创新能力。随着数字技术的不断发展，越来越多的企业开始意识到数字化转型的重要性，并加大在数字技术方面的投资。通过引入先进的数字技术和新的生产方式，对生产过程进行重塑，提升安全性和生产效率，以及开发新的产品。并且对商业模式也有一定的改进作用，可以增加管理效率，提高企业的核心竞争力。数字技术投资可以促进企业与外部环境的互动和连

接。数字技术的发展使得企业可以更加便捷地获取外部信息、资源和服务，从而更好地满足客户需求和市场变化。例如，通过数据分析技术，企业可以深入了解客户需求和行为习惯，为产品研发和服务提供更加精准的决策支持。同时，数字技术还可以帮助企业与供应商、合作伙伴等建立更加紧密的联系和合作，实现资源共享和互利共赢。

最后，数字技术投资可以提升企业的适应性和灵活性。随着市场环境的变化和技术的不断更新换代，企业需要不断地调整自身的战略和业务模式。数字技术可以帮助企业快速适应市场变化和技术创新，实现业务模式的灵活调整和创新发展。例如，云计算、大数据等技术的发展使得企业可以更加灵活地应对数据存储和分析的需求，提高企业的竞争力和市场地位。

(3) 本文发现当行业竞争强度和知识密集度均较高时，对企业数字化转型的贡献率可能为负，并且这种负贡献率可能会达到最大。这主要是因为较高的行业竞争强度和知识密集度可能会带来更多的不确定性和风险，使得企业在数字化转型过程中面临更大的挑战和压力。

一方面，较高的行业竞争强度可能会使得企业更加注重短期业绩和市场占有率，而忽视数字化转型的长远价值。在这种情况下，企业可能会缺乏足够的资源和精力投入到数字化转型中，导致数字化转型的进程受阻。

另一方面，知识密集度较高的行业往往要求企业具备更高的技术水平和创新能力，但也面临着更大的技术风险和市场不确定性。企业在进行数字化转型时，需要面对技术更新换代快、市场竞争激烈等挑战，如果不能及时跟上技术发展的步伐或者市场变化的需求，可能会导致数字化转型失败或者效果不佳。

## 5.2 管理启示

(1) 加大数字技术投资：企业应充分认识到数字技术投资对于数字化转型的关键作用。在数字化转型的过程中，企业需要增加对硬件设备、软件系统、数据安全等方面的投入。这包括但不限于云计算、大数据分析、人工智能、物联网等技术的应用。通过提供强大的技术支持，企业可以有效地提高数字化转型的效率和效果。

(2) 关注行业竞争强度和地区数字环境：企业在制定数字化转型战略时，应充分考虑行业竞争强度和地区数字环境的影响。不同行业的竞争状况和数字化需求有所不同，企业需要结合自身所在行业的竞争态势和地区数字化环境的特点，制定针对性的策略。例如，在竞争激烈的行业，企业可能需要更快速地进行数字化转型以保持竞争力；而在数字化环境较为落后的地区，企业可能需要更多地依赖于自身的数字化能力。

(3) 提升管理者能力：管理者在数字化转型中扮演着关键角色。企业应注重提升管理者的数字化思维、创新能力和战略规划能力。这可以通过培训、引进等方式实现。同时，企业还需要加强数字化转型的领导力和执行力，确保转型过程的顺利进行。

(4) 优化知识密集度：知识密集度对企业数字化转型具有重要影响。企业应注重提升技术人员的占比，加强员工培训和技能提升。此外，建立良好的知识共享和交流机制也是提高企业数字化转型效果的关键。这有助于促进不同部门、团队之间的合作与协作，推动数字化转型的深入进行。

(5) 制定全面转型策略：企业应制定全面的数字化转型策略，涵盖战略、组织、人才、技术等多个方面。这意味着企业在数字化转型过程中需要从顶层设计到具体执行各个层面进行全面规划。在战略层面，企业需要明确数字化转型的目标和路径；在组织层面，企业需要调整组织结构以适应数字化转型的需求；在人才层面，企业需要培养和引进具备数字化能力的人才；在技术层面，企业需要引入和开发适合企业需求的技术。

综上所述，本文的研究结论为企业在数字化转型过程中的决策和实践提供了重要的指导。充分考虑数字技术投资、行业竞争强度、地区数字环境、管理者能力以及知识密集度等因素，制定科学合理的策略和计划，有助于实现数字化转型的成功实施并获得长期竞争优势。

### 5.3 研究不足与展望

本文虽使用了多种模型和不同的影响因素，对企业数字化转型程度进行了研究，但仍有较多的不足。

(1) 本文的研究结论具有针对性，主要针对制造业产业。由于不同行业的数字化需求和转型路径存在差异，因此这些结论可能并不完全适用于其他行业。未来研究应考虑不同行业的特点，探讨数字化转型在不同行业中的适用性和差异性。

(2) 本文在研究数字化转型的影响因素时，可能未能涵盖所有相关因素。一些影响因素可能由于难以量化或者数据获取困难而未被纳入研究范围。为了更全面地理解数字化转型的影响机制，未来研究将尝试构建或寻找合理的代理指标，以补充和完善现有的因素框架。

为了克服这些局限性，未来的研究可采取以下措施：

(1) 扩展研究范围，包括不同行业的样本，以评估数字化转型在不同背景下的影响力和效果。

(2) 通过自行构建或寻找现有的代理指标，补充和完善影响因素的全面性，确保研究结果的准确性和适用性。

## 参考文献

- [1] BCG.2018BCG Global Challengers: Digital Leapfrogs[R].2019.
- [2] Murphy J T,Carmody P,Surborg B. Industrial Transformation or Business as Usual?Information and Communication Technologies and Africa’s Place in the Global Information Economy[J]. Review of African Political Economy,2014,41 (140): 264-283.
- [3] Dewan S,Kraemer K L. Information Technology and Productivity: Evidence from Country-Level Data[J]. Management Science,2000,46 (4): 548-562.
- [4] Zhang L,Wang Y. On the Path to Drive the Revival of Old Industrial Base in Northeast China by Informatization[J]. Proceedings of 2007 International Conference on Public Administration (3rd) ,Vol II,2007:487-494.
- [5] Saldivar A A F,Li Y,Chen W,et al. Industry 4.0 with Cyber-Physical Integration: A Design and Manufacture Perspective[J]. 2015 21st International Conference on Automation and Computing (ICAC) .2015:260-265.
- [6] Wollschlaeger M,Sauter T,Jasperneite J. The Future of Industrial Communication[J].IEEE Industrial Electronics Magazine.2017,11 (1): 17-27.
- [7] Gon P M. Horizontal Integration Between Cyber Physical System Based on Industry 4.0 and Manufacture Execution Systems through Middleware Building[J]. Journal of Korea Multimedia Society,2014,17 (12): 1484-1493.
- [8] Tang D. Industrial Network Integration Development and Upgrading Model Based on Industry 4.0[M]. Aer-Advances in Engineering Research,Yue X G,Cao Y,Habib M M,et al,2016: 43,70-74.
- [9] Repkine A. ICT Penetration and Aggregate Production Efficiency: Empirical Evidence for a Cross-Section of Fifty Countries[M],2008.
- [10]Rockart J F,Short J E. IT In The 1990s: Managing Organizational Interdependence[J].MIT Sloan Management Review,1990,30 (2): 7.
- [11]Clemons E K,Row M C. Electronic Consumer Interaction,Technology-Enabled Encroachment, and Channel Power : The Changing Balance Between

- Manufacturers Electronic Distribution and Established Retailers[M]:IEEE,1998:321-328.
- [12]Schreyer P. The Contribution of Information and Communication Technology to Output Growth[J]. STI Working paper,2000.
- [13]Sambamurthy V,Bharadwaj A,Grover V. Grove V. Shaping Agility Through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms[J]. MIS quarterly,2003:237-263.
- [14]Chi L,Ravichandran T,Andrevski G. Information Technology,Network Structure,and Competitive Action[J]. Information Systems Research,2010,21(3): 543-570.
- [15]Mač iulytė-Šniukienė A,Gaile-Sarkan E. Impact of Information and Telecommunication Technologies Development on Labour Productivity[J]. Provedia-Social and Behavioral Sciences,2014,110:1271-1282.
- [16]Zerenler M,Bilge F A,özilhan D. The Impact of Using Information Technologies on Crisis Management Success in Small and Medium Sized Enterprises[J]. The Business Review,2007,8(1):332-338.
- [17]Melián-González S , Bulchand-Gidumal J. A Model that Connects Information
- [18]hcnology and Hotel Performance[J]. Tourism Management,2016,53:30-37.
- [19]Parsa A B,Movahedi A,Taghipour H,et al.Toward Safer Highways,Application of XGBoost and SHAP for Real-Time Accident Detection and Feature Analysis[J].Accident Analysis&Prevention,2020,136:105405.
- [20]Mangalathu S,Hwang S H,Jeon J S.Failure Mode and Effects Analysis of RC Members Based on Machine-learning-basedSHapley Additive ex Planations (SHAP) Approach[J].Engineering Structures,2020,219:110927.
- [21]BAYUS B, ERICKSON G, JACOBSON R. The financial rewards of new product introductions in the personal computer industry[J]. Management Science, 2003,49(2):197-210.
- [22]Luthans F. THE CONTINGENCY THEORY OF MANAGEMENT: A path out of

- the jungle[J].Business Horizons, 1973, 16(3):67-72.
- [23] Joe, Tidd. Innovation management in context: environment, organization and performance[J].International Journal of Management Reviews, 2001.
- [24] VIAL G. Understanding digital transformation: a review and a research agenda[J]. The Journal of Strategic Information Systems, 2019, 28(2): 118-144.
- [25] TEECE D J, PISANO G, SHUEN A. Dynamic capabilities and strategic management [J] .Strategic Management Journal, 1997,18(7):509-533.
- [26] BRICKLEY J A, LEASE R C, SMITH JR C W. Ownership structure and voting on antitakeover amendments [J] .Journal of Financial Economics, 1988,20(1):267-291.
- [27] DEMERJIAN P R , LEV B , MCVAY S E. Quantifying managerial ability: a new measure and validity tests[J]. Management Science, 2011,58(7):1229-1248.
- [28] GHOSH D, OLSEN L. Environmental uncertainty and managers' use of discretionary accruals [J] .Accounting, Organizations and Society, 2009,34(2):188-205.
- [29] 陈醒. 说说美国先进制造业战略计划[J]. 国际融资, 2019(09):16-19.
- [30] 李金华. 中国先进制造业的发展现实与未来路径思考 [J]. 人文杂志, 2020(01):22-32.
- [31] 国家互联网信息办公室. 数字中国建设发展进程报告 (2019 年) [EB/OL]. [http://www.gov.cn/xinwen/2020-09/13/content\\_5543085.htm](http://www.gov.cn/xinwen/2020-09/13/content_5543085.htm)
- [32] 张培, 张苗苗. 制造企业数字化转型类型与触发机制 [J]. 管理现代化, 2020,40(06):19-24.
- [33] 马化腾. 数字经济[M]. 北京: 中信出版社. 2017.
- [34] 埃森哲. 2020 年技术展望 [EB/OL]. [https://www.accenture.com/\\_acnmedia/PDF-149/Accenture-Tech-Vision-2021-Full-Report-CN.pdf#zoom=40](https://www.accenture.com/_acnmedia/PDF-149/Accenture-Tech-Vision-2021-Full-Report-CN.pdf#zoom=40).
- [35] 埃森哲. 2020 中国企业数字转型指数研究 [EB/OL]. [https://www.accenture.com/\\_acnmedia/PDF-134/Accenture-China-Digital-Transformation-Index-2020.pdf#zoom=50](https://www.accenture.com/_acnmedia/PDF-134/Accenture-China-Digital-Transformation-Index-2020.pdf#zoom=50).

- [36] 吕铁.传统产业数字化转型的趋向与路径[J].学术前沿,2019(9):13-19.
- [37] 康鲲鹏,黄玉婷,董大维.试论如何推动电力行业数字化审计模式的发展[J].经济师,2019(12):128+130.
- [38] 国务院发展研究中心.传统产业数字化转型的模式和路径[R].2018.
- [39] 程叶青,王哲野,马靖. 中国区域创新的时空动态分析[J]. 地理学报,2014, 69 (12): 1779-1789.
- [40] 余江,孟庆时,张越,等. 数字创新:创新研究新视角的探索及启示[J]. 科学学研究,2017,35 (7): 1103-1111.
- [41] 李煜华,向子威,廖承军. 先进制造业数字化转型组态路径研究: 基于“技术 - 组织-环境”的理论框架[J]. 科技管理研究,2022,42 (03): 119-126.
- [42] 邢小强,周平录,张竹,等. 数字技术,BOP 商业模式创新与包容性市场构建 [J]. 管理世界,2019,35 (12): 116-136.
- [43] 沈运红,黄桁. 数字经济水平对制造业产业结构优化升级的影响研究基于浙江省 2008-2017 年面板数据[J]. 科技管理研究,2020,40 (3): 147-154.
- [44] 綦良群,王金石,崔月莹,等. 中国装备制造业服务化水平测度基于价值流动视角[J]. 科技进步与对策,2021,38 (14): 72-81.
- [45] 彭永涛,侯彦超,罗建强,等. 基于 TOE 框架的装备制造业与现代服务业融合组态研究[J]. 管理学报,2022,19 (3): 333-341.
- [46] 邢文杰,张景涛. 管理者能力与企业风险承担[J]. 会计之友,2022 (17): 79-87.
- [47] 陈雪琴,郑宝红. 国外管理者能力研究述评与展望[J]. 外国经济与管理,2018, 40 (7): 155-166.
- [48] 孙磊,张树山. 智慧化升级的驱动因素及其对物流企业绩效的影响[J]. 中国流通经济,2020,34 (2): 15-26.
- [49] 林艳,张欣婧.制造企业数字化转型不同阶段的影响因素: 基于扎根理论的多案例研究[J].中国科技论坛,2022(6):123-132,142.
- [50] 张路,李金彩,张瀚文,等. 管理者能力影响企业成本粘性吗?[J]. 会计研究,2019 (3): 71-77.

- [51] 吴先明, 苏志文. 将跨国并购作为技术追赶的杠杆: 动态能力视角[J]. 管理世界, 2014, 30 (4): 146-164.
- [52] 常嵘. 并购战略性新兴企业对传统企业转型升级的影响因素分析[J]. 经济理论与经济管理, 2017, 37 (12): 88-101.
- [53] 金珺, 李诗婧, 黄亮彬. 传统制造业企业数字化转型影响因素研究[J]. 创新科技, 2020, 20 (6): 22-34.
- [54] 贾建锋, 赵若男, 刘伟鹏. 数字经济下制造业国有企业转型升级的组态研究[J]. 研究与发展管理, 2022, 34 (2): 13-26.
- [55] 马中东, 宁朝山. 数字经济, 要素配置与制造业质量升级[J]. 经济体制改革, 2020 (3): 24-30.
- [56] 蒋盛益, 汪珊, 蔡余冲. 基于机器学习的上市公司财务预警模型的构建[J]. 统计与决策, 2010(9): 166-167.
- [57] 盛夏, 李斌, 张迪. 基于数据挖掘的上市公司信用评级变动预测[J]. 统计与决策, 2016(15): 159-162.
- [58] 黄志刚, 刘佳进, 林朝颖. 基于机器学习的上市公司财报舞弊识别前沿方法比较研究[J]. 系统科学与数学, 2020, 40(10): 1882-1900.
- [59] 佟金铎, 郭凤英, 翟兴, 等. 基于用户画像的患者就医影响因素研究[J]. 医学信息, 2021, 34(2): 11-14.
- [60] 孙发勤, 冯锐. 基于学习分析的在线学业成就影响因素研究[J]. 中国电化教育, 2019(3): 48-54.
- [61] 张云中, 秦艺源. 社会化标注系统标签质量影响因素研究: 基于随机森林算法[J]. 图书情报工作, 2019, 63(24): 119-126.
- [62] 朱振涛, 杜明阳, 刘颖. 共享电动汽车使用意愿的关键影响因素研究[J]. 江汉大学学报(社会科学版), 2020, 37(6): 91-102, 127.
- [63] 王志宁. 基于 XGBoost 的员工离职预测及特征分析模型[J]. 数字技术与应用, 2021, 39(3): 193-196.
- [64] 曹睿, 廖彬, 李敏, 等. 基于 XGBoost 的在线短租市场价格预测及特征分析模型[J]. 数据分析与知识发现, 2021, 5(6): 51-65.

- [65]武常岐,张昆贤,周欣雨等.数字化转型、竞争战略选择与企业高质量发展——基于机器学习与文本分析的证据[J].经济管理,2022,44(04):5-22.
- [66]王新光.管理者短视行为阻碍了企业数字化转型吗——基于文本分析和机器学习的经验证据[J].现代经济探讨,2022,No.486(06):103-113.
- [67]赵磊,赵纳晖,杜雨晴.超常增长与数字化转型——来自制造业上市公司的经验证据[J].山西财经大学学报,2023,45(03):82-97.
- [68]吴江,陈婷,龚艺巍等.企业数字化转型理论框架和研究展望[J].管理学报,2021,18(12):1871-1880.
- [69]赵艺婷,刘文昌.制造企业数字化转型驱动因素的组态效应研究[J/OL].科学与管理:1-18[2023-06-05].
- [70]朱永明,牛蓝霄,李婧.中国情境下数字化转型前因组态和提升路径分析[J].科技管理研究,2023,43(06):144-154.
- [71]申明浩,庞钰标,谭伟杰.道不同可相为谋: 高管团队异质性与企业数字化战略[J].南方金融,2023,No.557(01):50-64.
- [72]李华民,龙宏杰,吴非.异质性机构投资者与企业数字化转型 [J] .金融论坛,2021,26(11):37-46,56.
- [73]唐松,伍旭川,祝佳.数字金融与企业技术创新 : 结构特征、机制识别与金融监管下的效应差异 [J] . 管理世界 ,2020,36(5):9,52-66.
- [74]李雷,杨水利,陈娜.数字化转型的前因组态与绩效研究——来自中国制造业上市公司的经验证据[J].科技进步与对策,2023,40(16):32-41.
- [75]朱永明,牛蓝霄,李婧.中国情境下数字化转型前因组态和提升路径分析[J].科技管理研究,2023,43(06):144-154.
- [76]傅国华,王涛,楼润平等.数字化投资对制造企业突破式创新的影响研究[J].会计之友,2023,(24):76-83.
- [77]苏汝劫,常宇豪.经济新常态下多元化与企业绩效——基于动态能力的视角[J].宏观经济研究,2019,41(7):136-147.
- [78]陈庆江,王彦萌,万茂丰.企业数字化转型的同群效应及其影响因素研究 [J] . 管理学报 ,2021,18(5):653-663.

- [79]郭道燕,黄国良,张亮亮.高管财务经历、风险偏好与公司超速增长——来自中国经济“黄金期”的经验证据[J].山西财经大学学报,2016,38(10)
- [80]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(07):130-144+10.
- [81]王子阳,魏炜,朱武祥,等.商业模式视角下的天虹数字化转型路径探索[J].管理学报,2020,17(12):1739-1750.
- [82]钟熙,宋铁波,陈伟宏,等.高管从军经历与企业道德行为:基于企业商业腐败行为的实证研究 [J] . 南开经济研究 ,2021(2):201-224.
- [83]王军,朱杰,罗茜.中国数字经济发展水平及演变测度[J].数量经济技术经济研究,2021,38(07)
- [84]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(07)
- [85]方帅,程雨涵,刘知等.企业“一把手”和“二把手”的价值观差异与企业创新:基于机器学习的经验证据[J].中国软科学,2024(S1):203-222.
- [86]吴世农,陈智瑜.基于财务和非财务信息的债券违约预警模型研究——来自机器学习方法的经验证据[J].厦门大学学报(哲学社会科学版),2023,73(06):108-121.

## 附 录

本文使用代码如下。

1. 填补空缺值、缩尾处理及标准化的代码

```
import pandas as pd
from sklearn.impute import KNNImputer

# 读取 Excel 文件，替换'your_file.xlsx'为实际的文件路径
df = pd.read_excel(R"C:\Users\jiayingbin\Desktop\33.xlsx")

# 指定你想要填充缺失值的列名列表
columns_to_impute = ["管理者能力", "数字技术投资", "研发人员", "研发投入", "高管风险偏好", "知识密
集度", "环境不确定性", "地区数字环境"] # 添加你需要填充的其他列

# 使用 KNN 填充缺失值
knn_imputer = KNNImputer(n_neighbors=5) # 选择 K 值，可以根据情况调整

# 逐列填充缺失值
for column_to_impute in columns_to_impute:
    df[column_to_impute] = knn_imputer.fit_transform(df[[column_to_impute]])

# 保存填充后的数据框为新的 Excel 文件，替换'output_file_filled.xlsx'为实际的输出文件路径
df.to_excel('output_file_filled.xlsx', index=False)
import pandas as pd

# 读取 Excel 文件，替换'output_file_filled.xlsx'为实际的文件路径
df = pd.read_excel(R"C:\Users\jiayingbin\Desktop\填补好空缺值.xlsx")

# 列名，排除前两列
columns_to_trim = df.columns[2:]

# 指定缩尾的百分比
trim_percentage = 0.01 # 上下 1%

# 遍历需要缩尾的列
for column in columns_to_trim:
    # 判断列是否为数值型
    if pd.api.types.is_numeric_dtype(df[column]):
        lower_bound = df[column].quantile(trim_percentage)
        upper_bound = df[column].quantile(1 - trim_percentage)
```

```
# 缩尾处理
df[column] = df[column].clip(lower=lower_bound, upper=upper_bound)

# 输出处理后的数据框
print(df)

# 保存填充后的数据框为新的 Excel 文件，替换'output_file_filled.xlsx'为实际的输出文件路径
df.to_excel('缩尾处理.xlsx', index=False)

import pandas as pd
from sklearn.preprocessing import StandardScaler

# 读取 Excel 文件，替换'output_file_filled.xlsx'为实际的文件路径
df = pd.read_excel(R"C:\Users\jiayingbin\Desktop\缩尾处理.xlsx")

# 列名，排除第一列和最后一列
columns_to_standardize = df.columns[1:-1]

# 指定 01 分类变量的列名
binary_columns = ['01 分类变量 1', '01 分类变量 2'] # 替换为你的实际 01 分类变量列名

# 初始化标准化器
scaler = StandardScaler()

# 遍历需要标准化的列
for column in columns_to_standardize:
    # 判断列是否为 01 分类变量
    if column not in binary_columns:
        # 标准化处理
        df[column] = scaler.fit_transform(df[[column]])

# 保存标准化后的数据框为新的 Excel 文件，替换'标准化后.xlsx'为实际的输出文件路径
df.to_excel('标准化后.xlsx', index=False)
```

## 2. 模型评估

```
import pandas as pd
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LinearRegression, Lasso
from sklearn.svm import SVR
from sklearn.ensemble import RandomForestRegressor
import lightgbm as lgb
import xgboost as xgb
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
```

```
# 读取 Excel 文件，替换'标准化后.xlsx'为实际的文件路径
df = pd.read_excel(R"C:\Users\jiayingbin\Desktop\标准化后.xlsx")

# 划分特征和目标变量
y = df.iloc[:, 1] # 第二列为因变量
X = df.iloc[:, 2:-1] # 第三列到倒数第二列为自变量

# 划分训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 定义模型
models = {
    'Linear Regression': (LinearRegression(), {'fit_intercept': [True, False]}),
    'Lasso Regression': (Lasso(), {'alpha': [0.01, 0.1, 1, 10]}),
    'SVR': (SVR(), {'C': [0.1, 1, 10], 'kernel': ['linear', 'rbf']}),
    'Random Forest': (RandomForestRegressor(), {'n_estimators': [100,200,300], 'max_depth': [None, 10, 20]}),
    'LightGBM': (lgb.LGBMRegressor(), {'num_leaves': [100,200,300], 'learning_rate': [0.05, 0.1, 0.2]}),
    'XGBoost': (xgb.XGBRegressor(), {'max_depth': [5, 7,9], 'learning_rate': [0.05, 0.1, 0.2]})
}

# 训练和评估模型
for name, (model, param_grid) in models.items():
    # 创建网格搜索对象
    grid_search = GridSearchCV(model, param_grid, scoring='neg_mean_squared_error', cv=5, n_jobs=-1,
    verbose=1)

    # 执行网格搜索
    grid_search.fit(X_train, y_train)

    # 最佳超参数
    best_params = grid_search.best_params_

    # 最佳模型
    best_model = grid_search.best_estimator_

    # 在训练集上做预测
    y_train_pred = best_model.predict(X_train)

    # 在测试集上做预测
    y_test_pred = best_model.predict(X_test)

    # 训练集评估指标
```

```
rms_train = mean_squared_error(y_train, y_train_pred, squared=False)
mae_train = mean_absolute_error(y_train, y_train_pred)
r2_train = r2_score(y_train, y_train_pred)

# 测试集评估指标
rms_test = mean_squared_error(y_test, y_test_pred, squared=False)
mae_test = mean_absolute_error(y_test, y_test_pred)
r2_test = r2_score(y_test, y_test_pred)

# 显示评估结果
print(f"Model: {name}")
print(f"Best Parameters: {best_params}")

print("\nTraining Set:")
print(f"RMS: {rms_train:.4f}, MAE: {mae_train:.4f}, R^2: {r2_train:.4f}")

print("\nTesting Set:")
print(f"RMS: {rms_test:.4f}, MAE: {mae_test:.4f}, R^2: {r2_test:.4f}")
print("-" * 40)

3.特征重要性分析
# 导入所需的库
import numpy as np
import pandas as pd
import xgboost as xgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error
import shap

# 读取数据
data = pd.read_excel(R"C:\Users\jiayingbin\Desktop\标准化后.xlsx")

# 划分自变量和因变量
X = data.iloc[:, 2:] # 第三列到最后一列为自变量
y = data.iloc[:, 1] # 第二列为因变量

# 划分训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 创建 XGBoost 模型（可以在这切换模型）
xgb_model = xgb.XGBRegressor(n_estimators=100, random_state=42)

# 训练模型
```

```
xgb_model.fit(X_train, y_train)

# 预测
y_pred = xgb_model.predict(X_test)

# 评估模型性能
mae = mean_absolute_error(y_test, y_pred)
print(f'Mean Absolute Error: {mae}')

# 特征重要性
feature_importance = xgb_model.feature_importances_

# 打印特征重要性
print("\nXGBoost Feature Importance:")
for feature, importance in zip(X.columns, feature_importance):
    print(f'{feature}: {importance}')

# SHAP 值
explainer = shap.TreeExplainer(xgb_model)
shap_values = explainer.shap_values(X_test)

# Mean(|SHAP value|)
mean_abs_shap = np.mean(np.abs(shap_values), axis=0)
print("\nMean(|SHAP value|):")
for feature, value in zip(X.columns, mean_abs_shap):
    print(f'{feature}: {value}')

4.摘要图和部分依赖图
# 导入所需的库
import numpy as np
import pandas as pd
import lightgbm as lgb
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error
import shap
from shap import summary_plot

# 读取数据
data = pd.read_excel(R"C:\Users\jiayingbin\Desktop\标准化后.xlsx")

# 划分自变量和因变量
X = data.iloc[:, 2:] # 第三列到最后一列为自变量
y = data.iloc[:, 1] # 第二列为因变量
```

```
# 划分训练集和测试集
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# 创建 LightGBM 模型
lgb_model = lgb.LGBMRegressor(n_estimators=300, random_state=42)

# 训练模型
lgb_model.fit(X_train, y_train)

# 预测
y_pred = lgb_model.predict(X_train) # 使用训练集进行预测以生成 SHAP 值

# 计算 SHAP 值
explainer = shap.TreeExplainer(lgb_model)
shap_values = explainer.shap_values(X_train) # 使用训练集计算 SHAP 值

# 生成摘要图
summary_plot(shap_values, X_train, feature_names=data.columns[2:], show=True) # 使用训练集生成摘要图,
并显示图形
# 选择一个特定的变量
selected_feature = X.columns[6]

# 单个特征的部分依赖图
shap.dependence_plot(selected_feature, shap_values, X_test, feature_names=X.columns, show=True)
```

## 后 记

行文至此，我十余年的学生时代就正式落下帷幕。

还记得研究生开学时，关老师在双选会上讲的“不要做个精致的利己主义者，要有胸怀若谷、为国为民的情怀”，就深深打动了我。于是，我立刻联系了关辉国老师，成为了老师的最后一名学生。感谢关老师这几年对我的照顾，在学术上和人生道路的指导下，为我提供了非常多的帮助。其实说来惭愧，刚考上研究生时，我总觉得自己可以坐得住硬板凳，以后继续往博士深造，关老师也对我考博充满期待。但现在我写这份致谢的时候，已经坐在了公司的工位上，开启了新一轮的抓耳挠腮。

说真的，段家滩校区真的蛮小，除了常去的图书馆和食堂，似乎也没有多少承载我记忆的地方。在学校中，印象最深的，就是和老师同学一起抗击新冠疫情的时候了，真的很感谢当时我们学院老师的付出。出了校门，左手边的陈记牛肉面，右手边的大自然烤肉，这两个地方一个承载了我早上去上课之前的回忆，一个承载了我晚上与同学举杯痛饮的回忆。上的每一节课，听老师讲的每一句话，走的每一段路，仿佛就在昨天。我甚至现在还能回忆起当时收到研究生录取通知书的欢欣，而现在连研究生生涯都已然成为过去式。

感谢关辉国老师的教诲，感谢每一位为我上过课的工商管理学院的老师，感谢我的舍友李义振、李肖和既是我的舍友也是我的同事高浩宁，感谢我的同学对我的帮助，感谢所有在我人生道路上帮助过我的人。

以前常常会思考，未来我会做什么。时至今日我也仍在思考，仿佛在海上的旅人凝视着无法被预测的大海。以后做什么，怎么做，我似乎没有答案。但希望我能无畏困难，不断前行。志之所趋，无远弗届；志之所向，无坚不入。这世上的每个人都是被时代浪潮裹挟着的一粒尘埃，身作为一名普通人。我要成为一名优秀的普通人，热爱生活，积极工作，在不断的探寻中找到自己热爱并具有时代价值的事情。付真心，行真事，得真我！