

基于机器学习的汽车企业研发投入强度与财务绩效的影响研究

王友浩¹, 黄河^{1,2*}

¹上海理工大学管理学院, 上海

²上海理工大学智慧应急管理学院, 上海

收稿日期: 2025年2月24日; 录用日期: 2025年3月26日; 发布日期: 2025年4月3日

摘要

在全球碳中和与智能化浪潮驱动汽车产业深度转型的背景下, 研发投入已成为车企构建核心竞争力的关键。本文创新性地采用XGBoost机器学习模型, 突破传统研究方法的局限, 深入剖析了汽车企业研发投入强度对财务绩效的非线性影响。基于156家A股上市汽车公司2021年的数据分析, 并对XGBoost模型参数优化。研究发现研发人员数量、研发人员占比及研发投入营收比是驱动财务绩效的关键要素, 且研发团队规模与企业财务表现呈显著正相关。据此, 本文为汽车企业提出切实可行的建议, 包括加强研发团队建设、优化研发投入规模和提升研发资金效率。本研究不仅为汽车企业战略决策提供理论支撑, 也为政策制定者和投资者理解研发投入的价值、把握汽车产业发展趋势提供重要参考。

关键词

机器学习, 财务绩效, 投资

Research on the Impact of R&D Investment Intensity and Financial Performance of Automobile Enterprises Based on Machine Learning

Youhao Wang¹, He Huang^{1,2*}

¹Business School, University of Shanghai for Science and Technology, Shanghai

²School of Intelligent Emergency Management, University of Shanghai for Science and Technology, Shanghai

Received: Feb. 24th, 2025; accepted: Mar. 26th, 2025; published: Apr. 3rd, 2025

*通讯作者。

文章引用: 王友浩, 黄河. 基于机器学习的汽车企业研发投入强度与财务绩效的影响研究[J]. 运筹与模糊学, 2025, 15(2): 280-288. DOI: 10.12677/orf.2025.152083

Abstract

Against the backdrop of the global carbon neutrality and intelligentization wave driving a profound transformation in the automotive industry, research and development (R&D) investment has become a crucial strategic element for automotive enterprises to build core competitiveness. This paper innovatively employs the XGBoost machine learning model, breaking through the limitations of traditional research methods, to deeply analyze the non-linear impact of R&D investment intensity on the financial performance of automotive companies. Based on the data analysis of 156 A-share listed automotive companies in 2021, and through the optimization of XGBoost model parameters, the study finds that the number of R&D personnel, the proportion of R&D personnel, and the ratio of R&D expenditure to operating revenue are key factors driving financial performance, and that there is a significant positive correlation between the size of the R&D team and the company's financial performance. Accordingly, this paper proposes practical and feasible recommendations for automotive companies, including strengthening R&D team construction, optimizing the scale of R&D investment, and improving the efficiency of R&D fund utilization. This research not only provides theoretical support for strategic decision-making in automotive enterprises but also offers an important reference for policymakers and investors to understand the value of R&D investment and grasp the development trends of the automotive industry.

Keywords

Machine Learning, Financial Performance, Investment

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

1.1. 汽车产业转型下的研发投入挑战

在全球碳中和目标与智能化技术革命的驱动下,汽车产业正经历着由传统制造向“新四化”(电动化、网联化、智能化、共享化)的深度转型[1]。根据统计,全球头部车企研发投入强度(R&D)已攀升至营收的6.8%~11.2%,显著高于制造业平均水平。这种技术密集型特征使得研发投入成为企业构建核心竞争力的关键战略,但同时也带来了显著的财务风险[2]。长达3~5年的研发周期、复杂的技术不确定性以及专利转化效率等问题,使得研发投入与财务绩效间的非线性关系亟待深入探究[3]。特别是在我国规划强调汽车产业链自主可控的背景下,量化评估研发投入的经济效益具有重要的政策与实践意义[4]。

1.2. 机器学习破解传统研究瓶颈

既有文献在研发投入与财务绩效关系的研究中,普遍采用面板回归、DEA等计量方法,但存在三方面局限[5]:其一,线性假设难以捕捉研发投入的边际效益递减、协同效应等复杂机制;其二,静态分析框架无法反映汽车产业技术迭代的动态特征[6];其三,财务绩效多维度指标(如托宾Q值、专利引用收益)的交互影响尚未有效建模[7]。现在机器学习算法被普遍用于汽车产业的研究,为解决这些难题提供了新路径:XGBoost这一机器学习算法可识别研发投入强度与财务绩效关系之间的复杂关系,这种数据驱动的研究范式,为突破传统计量经济学的理论边界提供了可能。

1.3. XGBoost 赋能汽车产业研发投入精准评估

针对传统研究的局限性, 以及机器学习算法在复杂关系识别上的优势, XGBoost 模型为汽车产业研发投入的经济效益评估提供了更强大、更精准的分析工具。首先, XGBoost 能够更有效地捕捉研发投入与财务绩效之间的复杂非线性关系。与线性回归假设不同, XGBoost 基于梯度提升框架的集成学习模型, 能够自适应地拟合研发投入强度与各项财务绩效指标(如托宾 Q 值)之间复杂的曲线关系。与随机森林不同, XGBoost 不仅能够通过集成多个决策树来提升预测精度, 更重要的是, 它采用了梯度提升策略, 即每一棵新树都在拟合之前所有树的预测残差, 从而更有效地关注和解决预测的“错误”部分, 提升模型的整体预测能力和鲁棒性。

2. 算法介绍

2.1. 算法概述

XGBoost 是一种基于梯度提升(Gradient Boosting)方法的机器学习算法。它主要包含对目标函数的优化。XGBoost 使用“加法模型”来构建预测模型。也就是说, 它通过将多个弱模型(通常是决策树)逐步加到一起, 来提升预测性能。每个新加的模型都是针对前一个模型的错误(残差)进行训练的。这些决策树之间是相互独立的, 每棵树都是根据前一棵树的预测误差来调整的, 最终将多个树的结果加权平均, 形成最终的预测结果。

目标函数

损失函数是度量模型预测结果与实际目标之间差异的函数。对于回归问题, 常见的损失函数有平方误差(MSE):

为了避免过拟合, XGBoost 在每棵树的训练过程中加入了正则化项。正则化项用于控制模型复杂度, 惩罚树的深度和每个节点的权重。正则化项的形式为

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

其中, T 是树的叶子节点数, γ 是控制树复杂度的超参数, λ 是控制叶节点权重的正则化系数, w_j 是第 j 个叶子节点的权重。

XGBoost 采用的是基于 CART (Classification and Regression Trees)的决策树模型, 树的每个节点是一个特征的二元分裂。分裂的目标是使得树的每个叶子节点包含的样本尽可能纯净(即目标值尽量相同)。

每棵树的构建是在当前模型的负梯度方向上进行的。梯度提升的核心思想是, 模型的更新(新树)是根据负梯度来进行的, 通过学习负梯度(误差)来优化模型。

2.2. 训练过程

首先, XGBoost 会初始化一个常数预测值, 通常是所有训练样本目标值的平均值或中位数。每一步, XGBoost 会计算负梯度, 并构建一棵树来拟合这个负梯度。每棵新树都会基于负梯度来调整之前模型的误差。每次新建一棵树后, 会通过更新公式来更新预测值。最终的预测值是所有树的加权和。XGBoost 使用贪心算法来进行树的分裂选择, 即每次分裂都会选择一个最佳的特征和分裂点, 使得每次分裂后的信息增益最大化。信息增益可以用来度量当前特征如何帮助减少模型的误差。

2.2.1. 特征选择与正则化

XGBoost 会自动选择对分裂贡献最大的特征。它会对每个特征计算增益、频率、覆盖度等指标, 帮助决定哪些特征应该进入模型。XGBoost 的正则化能够有效防止过拟合。正则化项会对模型的复杂度进

行惩罚,特别是对树的深度和叶子节点的权重进行控制。

2.2.2. XGBoost 特征重要性得分的计算方法:

在 XGBoost 中,特征重要性是通过计算各个特征在决策树中所贡献的度量来评估的。主要有以下几种计算方式。增益是指某个特征在分裂过程中所带来的信息增益,表示该特征能够减少多少目标函数的误差。具体来说,它衡量的是使用特征进行分裂时,目标函数(如损失函数)减少的量。增益计算公式如下:

$$\text{Gain} = \frac{1}{2} \left(\frac{(G_L + G_R)^2}{H_L + H_R + \lambda} - \frac{G_L^2}{H_L + \lambda} - \frac{G_R^2}{H_R + \lambda} \right)$$

其中, G_L, G_R 是左右子树的梯度和, H_L, H_R 是左右子树的二阶梯度和, λ 是正则化参数。增益越大,表示该特征对模型的改进贡献越大。这个方法强调每个特征在模型训练过程中对减少误差的贡献。

覆盖度衡量的是特征在树的分裂中涉及到的样本数,通常是所有使用该特征进行分裂的样本权重的总和。覆盖度较大的特征,表示它在分裂中处理了更多的数据样本。

$$\text{Cover} = \sum_{i=1}^n w_i$$

其中, w_i 是样本的权重。覆盖度较大的特征可能具有较强的普遍性,因为它涉及到更多的训练样本,因此通常也被认为是较为重要的特征。频率表示特征在所有树的分裂过程中被使用的次数。频率越高的特征,表示它在模型训练中出现的次数越多,可能对模型的最终预测结果有更大的影响。频率较高的特征说明它在多个树的分裂中被多次使用,具有较大的预测贡献。通过增益、频率、覆盖度等指标,我们可以更好地理解哪些特征在模型中扮演了重要的角色。

3. 案例分析

3.1. 汽车行业发展概述

中国汽车行业上市公司群体伴随着中国汽车工业的蓬勃发展而壮大,其发展历程大致可划分为起步、快速增长和结构转型三个阶段。早期上汽集团和一汽轿车等为代表的企业率先登陆资本市场,为行业发展奠定了基础。随后,随着中国汽车市场对外开放和消费需求井喷,比亚迪、长城汽车等自主品牌以及众多零部件企业纷纷上市,上市公司数量快速增加,资本市场为行业注入了强劲动力。近年来,面对市场增速放缓和产业升级的挑战,新能源汽车产业的崛起成为新的增长点,宁德时代、蔚来汽车等新兴企业也相继上市,为行业带来了新的活力,目前中国汽车行业上市公司群体已形成数量庞大、类型多元、地域分布广泛的格局,成为中国汽车工业发展的中坚力量。

中国汽车行业上市公司群体呈现出国有企业与民营企业并存、传统车企与造车新势力共舞的局面,涵盖了整车制造、零部件供应、销售服务以及新能源汽车等多个细分领域。国有企业如上汽集团等历史悠久,规模庞大,而民营企业如比亚迪、长城汽车等则机制灵活,市场反应迅速,新兴造车势力则专注于技术创新和模式变革。在业绩表现上,传统燃油车企面临增长瓶颈和转型压力,业绩分化明显,而新能源汽车产业链上的公司则受益于行业红利,普遍业绩亮眼,市值也水涨船高,行业内上市公司整体呈现出业绩与估值分化的特征。

展望未来,中国汽车行业上市公司既面临着前所未有的发展机遇,也面临着激烈的市场竞争和技术变革的挑战。新能源汽车市场的爆发式增长、智能化网联化技术的加速渗透以及国家政策的大力支持,都为上市公司提供了广阔的发展空间。然而,市场竞争的日益激烈、技术迭代的加速、原材料价格的波动以及国际贸易环境的不确定性,也给上市公司带来了诸多挑战。投资者应密切关注行业发展趋势和企业转型动态,深入

分析企业基本面，理性评估投资机会与风险，方能把握中国汽车行业上市公司未来发展的投资机遇。
所以本研究以汽车行业上市公司为研究对象，研究其 2021 年数据。

3.2. 案例数据

本文选取了根据证监会 2012 年行业分类标准，剔除 ST 企业后，属于汽车行业的 156 家 A 股上市公司作为研究对象。研究时间为 2021 年。本文选取 2021 年作为研究时间，主要基于数据的完整性、行业背景及政策影响等因素的综合考虑。首先，2021 年的财务和经营数据已完整披露，确保数据的准确性和可靠性，同时减少了信息不对称和数据修正的可能性。其次，2021 年是全球新冠疫情持续影响的关键时期，汽车行业面临供应链波动、芯片短缺等挑战，研究这一时间点的数据可以分析企业在特殊环境下的应对策略及行业发展的调整方向。此外，2021 年是我国“十四五”规划的开局之年，新能源汽车、智能网联汽车等领域受到政策大力扶持，研究该年度数据有助于观察政策红利对行业创新和研发投入的具体影响。最后，选取 2021 年可以避免近期市场波动对长期趋势分析的干扰，使研究结果更加稳定和具备代表性。因此，2021 年作为研究时间具有重要的现实意义和研究价值。数据均来自 CSMAR 详细描述见表 1。

Table 1. Data description
表 1. 数据描述

名称	解释
研发人员数量	研发人员数量
研发投入金额	研发投入金额
研发投入占营业收入比重	研发投入占营业收入比例，财报直接披露
托宾 Q 值	市值/资产总计

3.3. 数据解释

研发人员数量反映了企业在人力资源方面对研发工作的重视程度。是企业研发投入强度的一个重要体现。较多的研发人员意味着企业有能力进行更多的研发项目，推动技术创新和产品升级。这有助于公司保持核心技术优势。

研发投入金额：研发投入金额是企业对研发活动的资金支持，体现了企业对研发工作的重视和投入力度。研发投入金额的增加通常表明企业致力于技术创新和产品开发。

研发投入占营业收入比重反映了研发投入在企业营业收入中的相对重要性。较高的比重意味着企业将较大比例的收入用于研发，是企业寻求长期发展和创新的策略表现。

托宾 Q 值是衡量企业市场价值与资产重置成本的比率，计算公式为公司市值除以资产总计。较高的托宾 Q 值意味着企业的市场价值相对较高，反映出投资者对企业未来增长潜力的乐观预期。托宾 Q 值可以用来衡量企业的成长性，较高的 Q 值可能表明企业具有较强的创新能力和市场竞争力。

在处理前的数据分布图中(见图 1)，存在明显的极端值和长尾现象，导致数据分布较为不均衡。首先，研发人员数量分布呈现长尾特征，绝大多数公司研发人员数量较少，少数公司则有极高的研发人员数量，造成分布的极度偏斜。研发人员数量占比分布呈现出一个相对对称的分布，但也表现出一定的离散性，表明公司之间在研发人员占比上差异较大。研发投入金额分布同样呈现出明显的右偏，少数公司投入的研发金额极高，而大多数公司则集中在较低的投入水平。研发投入占营业收入比率分布则显示出较为极端的值，大部分公司在这一比率上较低，只有少数公司表现出极高的比率。处理前的分布图受极端值影响较大，呈现出高度不均衡的分布形态，可能会对数据分析和建模造成影响。对其使用 1.5 倍 IQR 法去除极端值。

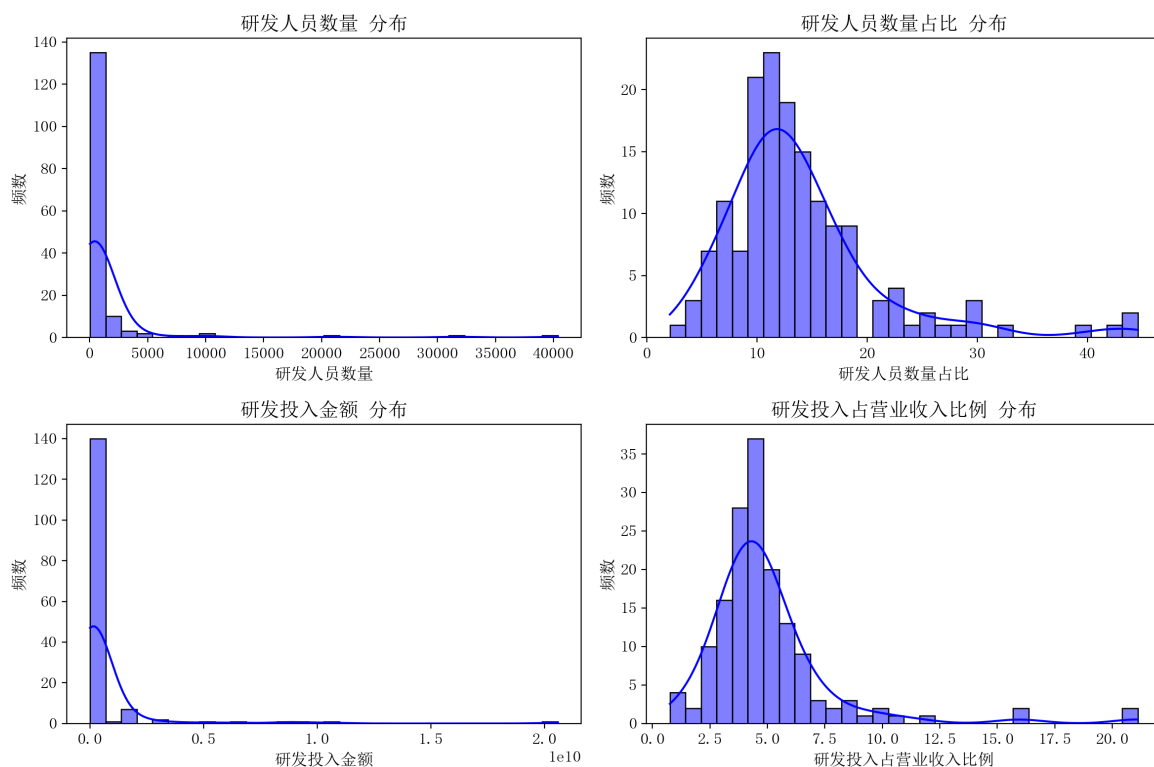


Figure 1. Distribution before data processing

图 1. 数据处理前分布图

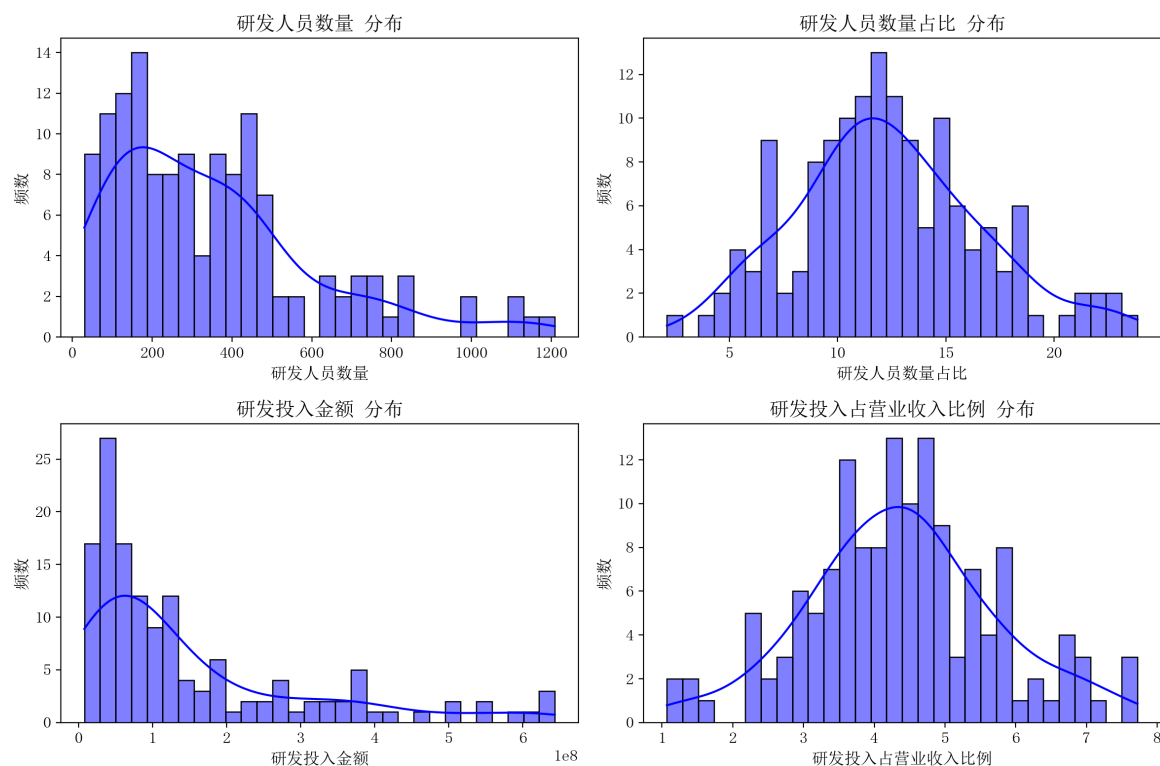


Figure 2. Distribution after data processing

图 2. 数据处理后分布图

经过去除极端值后的分布图显著改善了数据的呈现(见图 2)，变得更加合理和有代表性。首先，研发人员数量分布从原本的长尾分布转变为更加集中、具有轻微左偏的形态，显示大部分公司在研发人员数量上较少，只有少数公司拥有较多的研发人员。其次，研发人员数量占比分布趋近于正态分布，且大部分公司在研发人员占比上相对接近，这表明大多数公司在这方面保持在一个均衡的水平。研发投入金额分布则呈现出右偏趋势，反映出多数公司投入的研发资金较低，而极高的投入金额属于少数公司。最后，研发投入占营业收入比率分布显示出较为均匀的分布，并且呈现双峰结构，表明在研发投入占比上，不同公司之间存在较大差异。去除极端值后，数据的分布变得更加平滑且符合实际情况，这为后续的统计分析和建模提供了更好的数据基础。

3.4. 模型调优

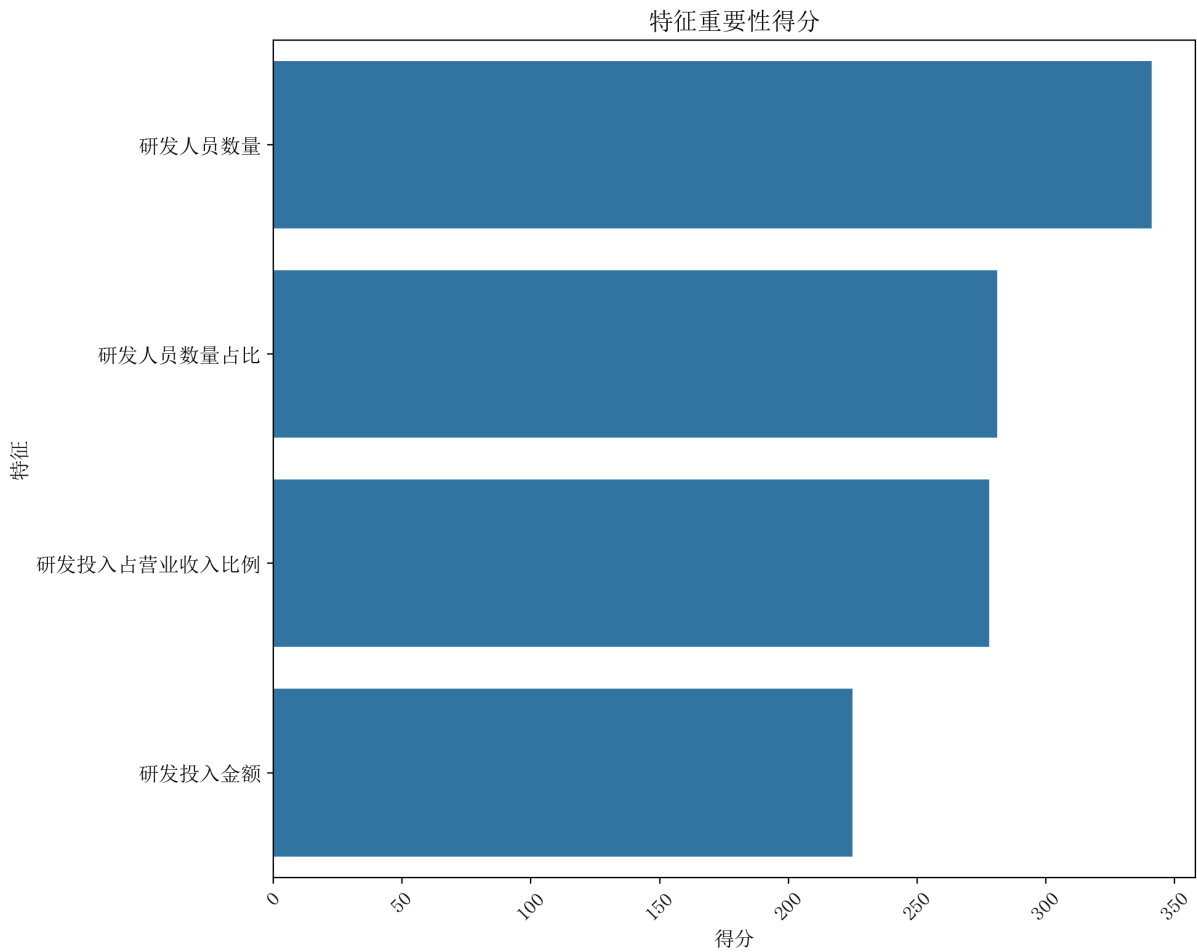


Figure 3. Feature importance ranking
图 3. 特征重要性排序

在 XGBoost 模型的调参过程中，通常会使用网格搜索(Grid Search)或随机搜索(Random Search)来寻找最优的参数组合。本研究使用网格搜索调参，因为它的参数搜索范围更全面。

max_depth (最大深度)参数范围是[3, 4, 5, 6]，它控制每棵树的最大深度。较大的深度可以使模型更复杂，但可能导致过拟合。Eta (学习率)参数范围是[0.01, 0.1, 0.2]，它控制每棵树的权重更新步长。较小的学习率可以提高模型的鲁棒性，但可能需要更多的树(更大的 n_estimators)。

Subsample (子采样比例)参数范围是[0.6, 0.8, 1.0], 它控制用于训练每棵树的样本比例。较小的比例可以减少过拟合, 但可能导致模型欠拟合。colsample_bytree (特征采样比例)参数范围是[0.6, 0.8, 1.0], 它控制在构建每棵树时使用的特征比例。较小的比例可以减少过拟合, 但可能导致模型欠拟合。

在本次调参中, 模型采用了以下参数配置: colsample_bytree 设置为 0.8, 表示在每次分裂节点时随机采样 80% 的特征, 有助于减少过拟合并提升泛化能力; eta 设置为 0.01, 较低的学习率使模型能够更细致地调整权重, 逐步逼近最优解, 提高精度, max_depth 设置为 5, 能捕捉数据中的复杂特征关系, 又能避免模型过于复杂导致的过拟合; subsample 设置为 1.0, 表示每次建树时使用全部样本数据, 确保模型充分利用样本信息。

使用最优参数后的模型结果见图 3 所示。

根据 XGBoost 回归模型的特征重要性排序结果, 研发人员数量在预测托宾 Q 值(财务绩效)中占据了最重要的位置, 得分最高。这表明, 企业的研发团队规模与其财务表现之间具有显著的相关性, 较多的研发人员通常意味着企业在创新和技术上的投入更为突出, 从而能够提升其市场竞争力和财务回报[8]。

排在第二位的是研发人员数量占比, 该特征显示出研发人员在公司整体员工中所占的比重。这一结果表明, 企业将更多资源投入到研发部门, 尤其是在人才方面, 可能有助于增强其创新能力和持续发展的潜力, 进而影响公司长期的财务表现和市场估值[9]。

相较而言, 研发投入占营业收入比率位居第三, 说明公司在营业收入中用于研发的比例对财务绩效有一定的影响。虽然这一特征的得分低于前两者, 但它仍然揭示了研发投入的相对重要性, 尤其是在行业内企业面临竞争压力时, 高研发投入比率可能代表了企业为维持竞争力所做出的战略选择[10]。最后, 研发投入金额排名最低, 表明单纯的研发资金投入对于托宾 Q 值的预测贡献相对较小, 可能是因为金额大小不能完全反映出企业在研发管理和创新效率上的能力[11]。

4. 结论建议

(1) 加大研发人才引进力度: 企业应积极拓宽研发人才引进渠道, 如与国内外知名高校建立长期合作关系, 设立联合实验室或实习基地, 定向培养符合企业需求的研发人才; 同时积极参与行业人才交流会和高端人才招聘会, 吸引具有丰富经验和创新能力的研发人才加入。此外, 还应利用网络招聘平台、猎头公司等多种渠道, 在全球范围内寻觅顶尖研发人才, 提升研发团队的整体实力。

(2) 构建多层次研发人才培养体系: 企业应建立完善的内部培训体系, 针对不同层级和专业方向的研发人员, 制定阶梯式的培训计划。可为新入职的研发人员提供基础技能和企业文化培训; 为资深研发人员提供专业技能提升和管理能力培训; 鼓励研发人员参加外部研讨会、技术交流论坛以拓宽视野, 学习行业前沿技术和动态。此外, 还可以推行“导师制”, 由经验丰富的资深研发人员指导年轻研发人员, 促进知识和技能的传承。

(3) 完善研发人才激励机制: 企业应建立与研发人员贡献和绩效挂钩的激励机制, 如实施股权激励、期权激励、项目分红等长期激励措施, 使研发人员的个人利益与企业发展紧密结合。设立奖项, 如设立研发创新奖、技术改进奖等荣誉性奖励, 对在研发工作中做出突出贡献的个人或团队给予表彰和奖励, 营造积极向上、鼓励创新的研发文化氛围。

(4) 优化研发团队组织结构: 企业应根据自身发展战略和研发项目特点, 构建灵活高效的研发团队组织结构。如推行项目型组织结构, 针对不同的研发项目组建跨部门、跨专业的研发团队, 充分发挥团队成员的协同效应; 建立扁平化的管理模式, 减少管理层级, 提高研发团队的沟通效率和决策效率; 鼓励研发团队内部进行技术交流和知识共享, 促进创新思想的碰撞和产生。

(5) 建立行业基准对标体系: 企业应密切关注国内外汽车行业研发投入强度的平均水平和领先企业

的投入强度,并结合自身所处的细分市场、发展阶段和技术战略,设定合理的研发投入强度目标。定期进行行业对标分析,根据市场变化和技术发展趋势,动态调整研发投入强度目标,确保研发投入的有效性。

(6) 紧盯研发强度和团队实力:重点关注研发投入高、研发团队强的汽车企业。研发投入占比高,说明企业重视技术创新;研发团队规模和质量是创新能力的关键保障,这些是未来竞争力的核心。

(7) 坚持长期投资视角:汽车产业的变革是长期的,研发投入的回报也需要时间才能体现。避免短期炒作,着眼于企业长期技术积累和市场潜力进行投资。因为汽车产业转型升级是一个长期过程,短期业绩波动不代表长期价值。

(8) 关注政策:关注符合政策支持方向(如新能源、智能化)且积极与产业链上下游协同发展的企业。政策支持带来红利,产业协同提升竞争力。

基金项目

教育部人文社会科学项目,数字经济赋能企业供应链韧性的影响效应、作用机制与提升路径研究,23YJC790046。

参考文献

- [1] 周蔡逸, 巩嘉荟, 李若萱, 等. ESG 视角下的新能源汽车商业模式与财务绩效研究——以小鹏汽车为例[J]. 商场现代化, 2025(3): 177-179.
- [2] 朱瑾仪. 企业股权激励措施对财务绩效的影响——以蔚来汽车为例[J]. 中国市场, 2024(36): 145-148.
- [3] 李磊. 领先型企业的研发策略对汽车制造业上市公司的绩效影响分析——基于上汽集团的案例研究[J]. 江苏商论, 2024(12): 102-105.
- [4] 夏文蕾, 余辉, 程佳银, 等. 数字化背景下企业环境绩效影响因素组态与路径研究——以新能源汽车制造业为例[J]. 财会通讯, 2024(16): 89-95.
- [5] 柳德才, 李冰清. “双碳”目标下新能源汽车上市公司财务绩效评价研究[J]. 生产力研究, 2024(4): 149-154.
- [6] 梁积江, 刘文亮. 新能源汽车产业创新能力如何影响财务绩效——基于中国上市公司的实证分析[J]. 产业经济评论, 2023(6): 5-26.
- [7] 吴文劲. 绿色转型升级背景下湖北新能源汽车产业链发展分析[J]. 湖北社会科学, 2023(6): 66-72.
- [8] 宁怡. 新冠疫情背景下汽车制造上市公司财务绩效评价——基于因子分析法[J]. 商场现代化, 2022(20): 174-176.
- [9] 王玥晗. 研发投入、绿色技术创新对财务绩效的影响研究[J]. 会计师, 2022(19): 1-4.
- [10] 乔安琪, 王海燕, 聂昕然. 政府补贴对新能源汽车企业财务绩效影响研究——以比亚迪公司为例[J]. 唐山师范学院学报, 2022, 44(4): 104-108.
- [11] 卓敏, 陈丽丽, 朱家明. 基于 A 股汽车制造业对所有权结构、研发投入与企业绩效的研究[J]. 黑龙江大学自然科学学报, 2022, 39(3): 300-306.