

员工晋升影响因素分析及预测研究

韩是坚, 罗鄂湘, 田嘉豪

上海理工大学管理学院, 上海

收稿日期: 2025年7月27日; 录用日期: 2025年8月19日; 发布日期: 2025年8月27日

摘要

本文以Kaggle平台提供的大型跨国公司员工晋升数据集为样本, 应用多种机器学习算法构建晋升预测模型, 确定影响员工晋升的重要因素。研究通过特征工程优化数据, 在经过SMOTE进行数据平衡后, 应用Logistic回归、决策树、随机森林、XGBoost和Stacking五种算法建模, 结果显示XGBoost模型预测准确率最高(94.19%)。研究还经过SHAP分析揭示了培训评估平均分、部门、总绩效表现、上一年培训次数和服务年限是影响晋升的关键前五因素。本研究为企业人力资源管理选择预测算法和优化晋升体系提供了数据支持与针对性建议。

关键词

员工晋升, 机器学习, 特征选择, SHAP

Analysis and Prediction of Influencing Factors of Employee Promotion

Shijian Han, Exiang Luo, Jiahao Tian

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

Received: Jul. 27th, 2025; accepted: Aug. 19th, 2025; published: Aug. 27th, 2025

Abstract

Based on the employee promotion dataset of large multinational companies provided by the Kaggle platform, this paper uses a variety of machine learning algorithms to construct a promotion prediction model and identify the important factors affecting employee promotion. In this study, the data is optimized by feature engineering, and the data is balanced by SMOTE, and five algorithms are used to model the data: logistic regression, decision tree, random forest, XGBoost and Stacking, and the results show that the XGBoost model has the highest prediction accuracy (94.19%). The study also revealed that the average_training_score, department, sum_metric, no_of_trainings, and

length_of_service were the top five key factors influencing promotion. This study provides data support and targeted suggestions for the selection of prediction algorithms and the optimization of promotion system for enterprise human resource management.

Keywords

Employee Promotion, Machine Learning, Feature Selection, SHAP

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

在动态竞争的商业环境中，完善的人才储备机制是企业可持续竞争力的核心支柱。政策层面，《中共中央关于进一步全面深化改革、推进中国式现代化的决定》明确提出需强化人才激励机制，凸显人才管理的战略地位。有效的晋升机制是人才激励机制中的一个重要策略。然而，传统晋升机制存在显著缺陷：一方面，传统晋升机制过度依赖历史绩效而忽视岗位适配性，如 Fernandez-Lozano 等(2019)证实灵活工作安排者晋升机会系统性降低[1]；另一方面，制度模糊性引发公平性质疑，进而削弱员工积极性，唐乐等(2019)指出政治技能通过上下级关系中介效应提升晋升概率[2]，王庆娟等(2023)发现晋升公平会引发员工的情绪变化和认知变化，进而改变员工 - 组织关系、工作态度及工作动机[3]，而苏屹等(2018)研究发现晋升制度模糊化对员工离职倾向有显著的正向影响[4]。

鉴于人力资源管理问题的复杂性，需要进行多理论协同效应的验证。人力资本理论强调员工通过教育、培训、工作经验积累的人力资本是职业发展的核心驱动因素，员工的服务年限、培训次数等作为人力资本积累的关键指标；而信号理论则指出培训评估成绩、绩效表现等信号会向组织传递员工能力信息，进而影响晋升决策，而部门差异可能反映不同信号的权重差异。

机器学习的发展为多理论协同效应验证提供了技术支持。随着机器学习在企业人力资源管理领域的研究发展，为系统识别晋升影响因素、建立更精确的员工晋升预测模型提供了研究基础。张敏等(2022)回顾人工智能在人力资源管理领域的研究发现，机器学习在战略规划、招聘、培训、绩效管理、薪酬制度、员工关系管理等领域有着广泛应用[5]。Raza 等(2022)利用扩展树分类器(ETC 算法)实现 93%预测精度[6]；王联英等(2021)通过决策树算法的人力资源推荐，提升招聘平台的人力资源推荐质量[7]。人工智能在 HRM 的应用虽广泛，却面临四大挑战：HR 现象的复杂性、小数据约束、算法公平性争议及员工对数据化管理的抵触[8]，持续性数据训练显得尤为重要。

机器学习在员工晋升预测模型建立方面的研究已有一定的进展。Muhannad I 等(2023)发现通过机器学习算法，如逻辑回归、支持向量机和随机森林，可以有效预测和辅助员工晋升决策，提升晋升过程的效率和准确性[9]。Amir-Mohsen K 等(2015)提出了一种基于强化学习的方法来优化不同知识水平员工的晋升策略，以提高生产效率和利润[10]。Zhang 等(2020)基于人口统计学特征，采用 5 种机器学习算法预测晋升，其中随机森林准确率最高(83.4%)，证实人口特征与晋升强相关[11]。Jafor 等(2023)通过改进 AdaBoost (以 ANN 为基分类器)，在 Kaggle 的员工晋升评估数据集上实现最优综合性能(准确率 95.3%，F1 值 95.76%)，优于单一算法[12]。

为系统识别员工晋升影响因素、持续增加员工晋升预测模型训练经验，本文基于具有 54,808 条样本的 Kaggle 跨国公司数据集，对比 Logistic 回归、决策树及集成算法(XGBoost/Stacking/随机森林)的预测效

能。本研究旨在验证培训、服务年限等人力资本要素对晋升的驱动作用，以及绩效、培训评估等信号对晋升决策的影响机制在跨国企业情境中的适用性，并通过机器学习方法补充多维度因素交互作用的识别，目标在于构建适应中国企业管理情境的高精度晋升预测模型；通过特征重要性分析识别关键晋升影响因素；提出兼顾算法效率与组织公平性的晋升机制优化路径，为企业人才配置决策提供数据驱动支持。

为此，本文使用了 Python 语言作为分析和建模的工具，建模流程如图 1 所示。

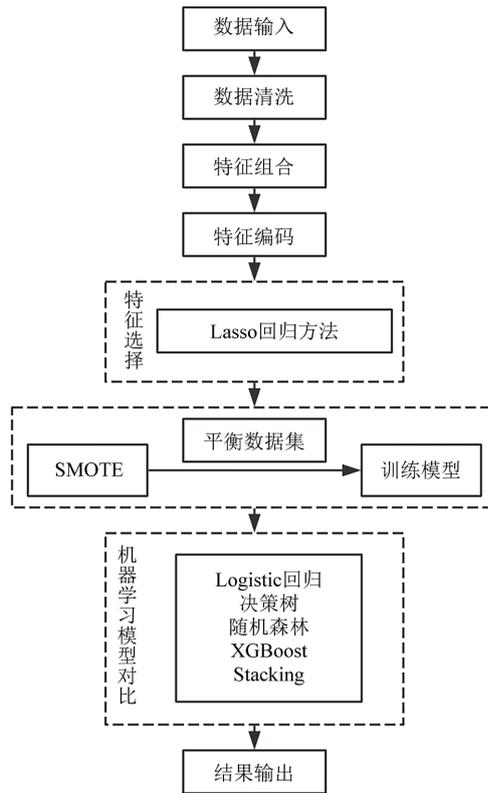


Figure 1. Modeling process
图 1. 建模流程图

2. 数据简介与预处理

本节在对本研究所用数据集进行描述后，应用数据清洗、特征组合、特征编码进行数据预处理。

2.1. 数据简介

本文所用数据来自 Kaggle 平台发布的员工晋升数据集。该数据集包含 9 个广泛垂直领域的大型跨国公司的真实员工数据，共有 54,808 个员工样本，12 个特征变量和 1 个目标变量，包括员工的基本信息(员工编号、部门、年龄、学历等)与影响员工晋升的各种因素(培训次数、工龄、考核分数等)，如下表 1 所示。由于数据来自实际场景，数据集保证了样本的真实性和客观性。

Table 1. Dataset field description
表 1. 数据集字段说明

字段	含义	类型	取值范围
employee_id	员工编号	名义	1~78,298

续表

department	所在部门	名义	Analytics/Finance/HR/Legal/ Operations/Procurement/R&D/Sales & Market- ing/Technology
region	就业地区	名义	1~34
education	学历	定序	Bachelor's/Below Secondary/Master's & above
gender	性别	名义	f/m
recruitment_channel	招聘渠道	名义	other/referred/sourcing
no_of_trainings	上一年培训次数	离散	1~10
age	年龄	连续	20~60
previous_year_rating	上一年评级	定序	1~5
length_of_service	服务年限	离散	1~37
awards_won?	上一年是否赢得奖项	二分类	0 = 否, 1 = 是
avg_training_score	培训评估的平均得分	连续	39~99
is_promoted	是否推荐晋升	二分类	0 = 否, 1 = 是

2.2. 数据清洗

数据清洗指识别并修正数据集中的缺失值、异常值和重复记录等问题的过程。其核心目标是提升数据的准确性、完整性和一致性，为后续数据分析与建模奠定可靠基础。本文涉及的数据清洗工作主要包含以下关键步骤。

不相关的变量处理：通过了解变量含义发现，员工编号字段 `employee_id` 对于研究员工晋升问题无实际意义，为无关变量，因此删去对应的列，剩余 11 个特征变量与 1 个目标变量进行探索性数据分析。

缺失值处理：学历缺失 2409 条、上一年评级缺失 4124 条。针对缺失机制分析：上一年评级缺失样本的服务年限均为 1 (表明为新入职员工)，故以 0 值填充(代表无历史评级)；学历作为类别变量，因本科学历占比显著高于其他类别，采用众数填充。

异常值处理：上一年培训数量及员工年龄的异常值具有实际业务意义(如反映特殊人才/高龄专家)，且样本量有限，故予保留；服务年限存在若干高于上限的异常值(如>13 年)，采用缩尾处理(Winsorization)将其限制为最大值 13 年。

数据清洗后剩余 54,490 个样本和 12 个特征。

2.3. 特征组合

鉴于单一特征难以全面捕捉员工表现的复杂模式，本文通过特征组合构建交互指标以增强综合评估能力。具体新增特征如下：

① 将绩效相关的两个变量，上一年是否赢得奖项 `awards_won?`和上一年评级 `previous_year_rating` 相加得到一个代表总绩效表现的指标 `sum_metric`。该设计解决了单一特征无法反映“高奖项低评级”类员工特殊贡献的问题。

② 将培训相关的两个变量，培训评估的平均得分 `avg_training_score` 和上一年培训次数 `no_of_trainings` 相乘得到一个代表总培训表现的指标 `total_score`。此组合可以更好地反映出在培训中的整体投入和效果。

特征工程后新增 2 个组合特征，当前特征总量为 14。组合特征通过协同效应增强模型对“特殊贡献员工”和“培训效能异常员工”的识别能力。

2.4. 特征编码

为消除类别变量对机器学习模型的输入障碍，本文通过编码转换提升数据可计算性，确保模型的输入一致性，提高模型的训练速度。具体实施如下：

① 对具明确等级关系的学历(硕士及以上 > 本科 > 中学及以下)，采用标签编码(Label Encoding)映射为有序数值(2/1/0)，确保学历层级信息被完整保留。

② 对无内在顺序的部门、性别、招聘渠道，通过独热编码(One-Hot Encoding)生成二进制特征列。该方法可消除类别间伪序风险，同时增强模型对独立类别效应的捕捉能力。

3. 基于 Lasso 的特征选择

在进行数据预处理后，此时数据集共有 20 个特征。较多的特征值不仅会增加计算复杂度，也会带来过拟合风险，为了避免这种情况，本文采用 Lasso 回归方法进行特征筛选，以提高模型的处理效率。

Lasso 回归(Lasso Regression)是一种线性回归的正则化形式，通过引入 L1 范数惩罚项来实现模型的稀疏性，从而有助于特征选择。通过 L1 正则化强度参数 C 与特征系数的动态映射关系，绘制了 Lasso 正则化路径图，如图 2 所示。从图 2 中可以看到：当正则化强度极大时，特征系数均被压缩至零轴基线。随着正则化约束逐步放松，更多特征系数从零开始增长。Lasso 特征选择在参数 $C = 0.05$ 时，通过 L1 正则化从数据中筛选出包括 education、age、department 等在内的 17 个特征值。

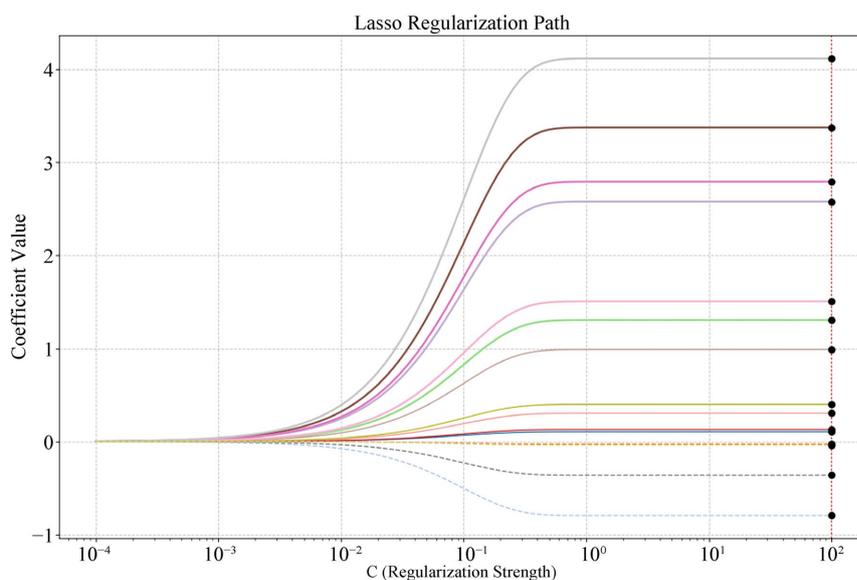


Figure 2. Lasso regularization path
图 2. Lasso 正则化路径

4. 员工晋升预测

4.1. 不平衡数据处理

经过数据清洗和特征工程后发现在 50,124 条样本中，未晋升员工占比为 90.77%，远大于晋升员工占比的 9.23%。因此，本数据集为不平衡数据集，这会导致模型在训练过程中更多地受到多数样本的影响，

使模型在少数类别上表现不佳，所以需要进行平衡处理。

本研究采用分级处理策略：首先通过 `train_test_split` 函数以 7:3 比例划分数据集，随后对训练集施加 SMOTE 过采样技术，最终获得训练样本 63,724 个，其中晋升与未晋升样本各 31,862 个，为后续建模提供无偏数据基础。

4.2. 分类器对比分析

本文应用 Logistic 回归、决策树、随机森林、XGBoost 和 Stacking 算法训练模型。

4.2.1. 机器学习算法原理

Logistic 回归通过逻辑函数(Sigmoid)将线性组合映射为[0, 1]概率值，专门解决二分类问题。其本质是分析自变量与因变量间的非线性关系，输出结果具有明确概率解释。决策树(Decision Tree)是一种广泛应用于分类和回归任务的机器学习算法，采用树形递归分割机制，决策树通过特征阈值划分将数据空间转化为层次化决策路径，最终在叶子节点输出预测结果。随机森林(Random Forest)是一种基于决策树的 Bagging 集成学习算法，通过多数投票机制聚合弱分类器预测结果，显著提升模型鲁棒性。XGBoost (eXtreme Gradient Boosting)是在 GBDT (Gradient Boosting Decision Tree)的基础上改进，基于决策树的 Boosting 集成学习算法。XGBoost 通过加法模型迭代训练 CART 弱学习器，每一轮聚焦修正前序残差，并引入正则化项控制复杂度。Stacking 算法区别于同质集成的 Boosting/Bagging，Stacking 属于异质集成方法，它通常是基于多个不同学习器的集成，也被称为异质集成方法。Stacking 模型一般为双层结构，第一层用研究中表现最好的三个模型为基分类器，第二层使用 logistic regression 作为元分类器。

4.2.2. 各模型调参与结果比较

在超参数优化过程中，研究选用了随机搜索(Random Search)方法，以高效探索复杂参数空间并提高模型性能。随机搜索方法通过在预设的参数范围内进行概率抽样，有效突破了传统网格搜索全组合遍历的局限性，特别适用于高维参数场景或在计算资源有限的情况下。借助随机搜索方法，各模型的调参过程及最终确定的最优参数如下表 2 所示。

Table 2. The hyperparameter tuning process and results of each model

表 2. 各模型调参过程与结果

模型	关键参数	取值范围	最佳值
Logistic 回归	C	[0.01, 0.1, 1, 10, 100]	100
	penalty	['l1', 'l2']	l2
	solver	['liblinear']	liblinear
	criterion	['gini']	gini
决策树	max_depth	['10', '12', '15']	15
	min_samples_leaf	['1', '2', '3']	1
	min_samples_split	['5', '6', '7']	6
	criterion	['entropy']	entropy
随机森林	max_depth	['10', '12', '15']	15
	min_samples_leaf	['1', '2', '3']	1
	min_samples_split	['5', '6', '7']	5
	n_estimators	['100', '200', '300']	200

续表

XGBoost	learning_rate	['0.01', '0.03', '0.06', '0.08']	0.08
	max_depth	['10', '12', '15']	15
	subsample	['0.7', '0.8', '0.9', '1']	1
	colsample_bytree	['0.5', '0.7', '0.9']	0.9
	gamma	['0', '0.1', '0.2', '0.3']	0
Stacking			0.8986

将五种分类器的预测精确度、F1-Score 和 AUC 值进行汇总比较，结果如表 3 所示。同时绘制了如图 3 所示的 ROC 曲线。

Table 3. Performance of different models
表 3. 不同模型表现

分类器	Accuracy	Precision	Recall	F1-Score	AUC
Logistic 回归	80.40%	24.03%	51.95%	32.85%	0.7440
决策树	89.99%	46.15%	50.94%	48.42%	0.8364
随机森林	91.38%	53.48%	50.43%	51.91%	0.8620
XGBoost	94.19%	76.72%	53.17%	62.81%	0.9015
Stacking	93.74%	70.13%	55.98%	62.26%	0.8986

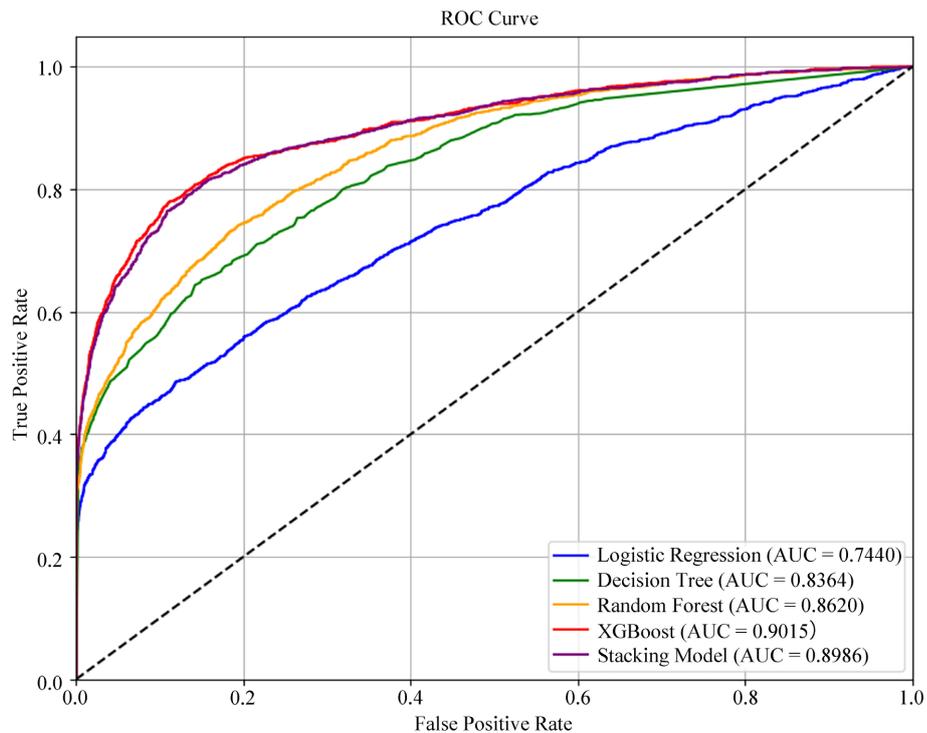


Figure 3. ROC curves for different classifiers based on SMOTE
图 3. 基于 SMOTE 的不同分类器 ROC 曲线

根据表 3 和图 3 中的信息，我们可以得出以下结论。

首先，从准确率(Accuracy)表现看，逻辑回归和决策树模型效果一般，其余三个模型均超过 90%，其中 XGBoost 模型以 94.19%的准确率表现最优。其次，在精确率(Precision)方面，XGBoost 模型同样领先，达到 76.72%，表明其识别正类样本的能力较强；Stacking 模型次之(70.13%)，逻辑回归模型表现不佳(24.03%)。接着，比较召回率(Recall)得分，Stacking 模型表现突出(55.98%)，XGBoost 模型次之(53.17%)，其余几个模型都在 51%上下。然后，在综合平衡精确率与召回率的 F1 分数上，XGBoost 模型以 62.81%的得分和 Stacking (62.26%)领先其他三个模型，证明了优越的综合能力。最后，为直观比较 AUC 值差异，图 3 的 ROC 曲线显示 XGBoost 模型表现最佳，AUC 值为 0.9015。

综合来看，随着模型复杂度增加，性能指标逐步提升：逻辑回归作为基线模型在处理复杂数据时表现有限；决策树通过非线性建模显著提升性能；随机森林和 XGBoost 通过集成学习进一步优化了鲁棒性和泛化能力。

5. 员工晋升影响因素分析

5.1. 员工晋升影响因素研究发展

晋升是指员工从现有职位向更高层次的职位变动，这一过程意味着员工需要承担更大的职责和风险。与此同时，也意味着员工将获得更大的职权和更优厚的待遇[11]。

实证研究专注于分析某些特定因素对员工晋升的影响。张辉华(2012)基于两家企业员工问卷调查，证实他评情绪智力与晋升显著正相关[13]。周文霞等(2015)通过 76 篇文献的元分析发现，人力资本、社会资本与心理资本均能正向预测主客观职业成功[14]。王建(2021)采用事件史分析法处理“社会网络与职业经历(JSNET)”并追踪调查数据，发现工作组织流动显著提升劳动者晋升机会[15]。张瑞娟等(2022)整合多视角文献指出：隐含偏见、工作-家庭支持缺失及领导认同匮乏构成女性晋升核心障碍[16]。吕芳(2020)同样发现中国女性领导干部晋升过程中要面临种种阻碍[17]。Ibarra H 等(2010)的研究中也发现了同样的结果[18]。Bode 等(2022)基于性别角色理论和组织承诺理论，通过纵向准实验设计和情景实验，发现参与企业社会责任项目可能会降低员工的晋升机会[19]。马玥等(2021)利用央企高管数据的非线性回归表明：员工业绩与晋升概率正相关，且薪酬差异或晋升概率越大时，业绩对晋升的影响敏感性越强[20]。

除此之外还有一些研究是以机器学习来探究影响员工晋升的因素。Zhang 等(2020)通过机器学习算法发现除了性别，不同的教育水平、年龄及服务年限都会对晋升产生影响[11]。Muhannad I 等(2023)研究发现培训次数、服务年限、上一年评级、年龄、获奖情况对员工晋升有着显著影响[9]。Jafor 等(2023)发现常见的关键影响因素有：生产效率、经验、种族等[12]。

综合来看，对于影响员工晋升的因素可归纳为以下几个方面。在影响员工晋升的个人特质中，情绪智力、政治技能、职业胜任力(包含人力资本、社会资本和心理资本)以及业绩表现都起着显著作用。具体而言，更高的情绪智力、更出色的政治技能和业绩表现，往往能提升获得晋升的机会。同时，通过机器学习算法还发现，除了性别，不同的教育水平、年龄及服务年限也会对晋升产生影响，培训次数、上一年评级、获奖情况以及生产效率、经验、种族等也是影响员工晋升的关键因素。此外，平衡工作与家庭生活同样至关重要，但存在“灵活性耻辱”现象：因照顾家庭而减少工作时间会限制职业发展前景，女性员工还面临着性别偏见以及缺乏工作-家庭支持文化等额外障碍。从组织与社会责任角度看，参与企业社会责任项目可能对男性的晋升产生不利影响，而组织内部的工作流动则有助于员工争取晋升机会。

5.2. 基于 SHAP 的员工晋升影响因素分析

SHAP 分析作为一种用于解释模型预测结果的方法，在提供了对模型决策过程的解释性同时，也

有助于改善模型的可解释性和可靠性。通过计算特征对预测结果的贡献程度来帮助理解模型的预测逻辑。

本文基于 SHAP 值计算特征重要性，如图 4 所示。

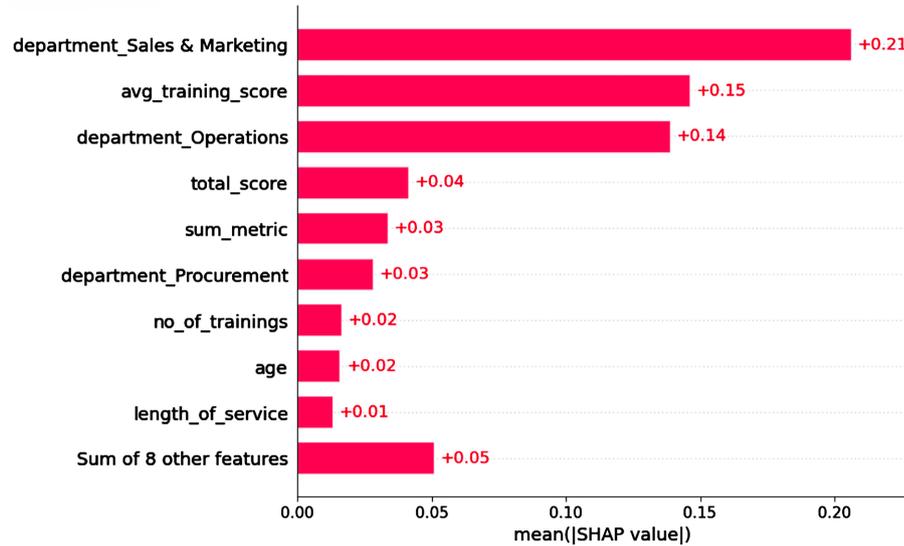


Figure 4. Feature importance ranking

图 4. 特征重要性排序

并进一步通过 SHAP 值蜂群图分析特征，见图 5。

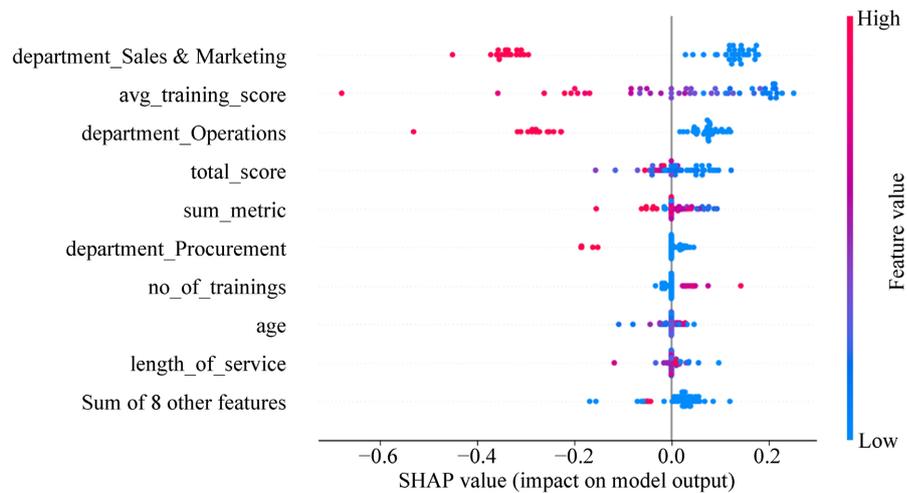


Figure 5. Feature SHAP values beeswarm plot

图 5. 特征 SHAP 值蜂群图

蜂群图直观展示了各特征对模型输出的具体影响(SHAP 值)及其特征值分布。图中每个点代表一个样本的 SHAP 值，其颜色反映该样本的特征值高低(红色高，蓝色低)。avg_training_score 为核心驱动因素，SHAP 均值达 0.15，且蜂群图中高分样本(红点)密集分布于高 SHAP 区(>0.1)，低分样本(蓝点)集中于负向区，表明其强正向影响。类似地，department、sum_metric 等高特征值样本也呈现 SHAP 值增大的趋势。no_of_trainings、length_of_service 高值样本轻微右偏(SHAP > 0)，暗示微弱正向关联。

5.3. 员工晋升因素交叉分析

在将通过 SHAP 分析得出的重要变量与因变量做交叉分析，结合交叉概率图进一步发现各变量与员工晋升的关系。

(1) 员工的培训评估平均分

图 6 是员工培训评估平均分与晋升情况的交叉图。根据图 6 数据可得：员工成绩集中于中等区间(峰值约 60~80 分)，两端(低分<40、高分>90)占比较低。这反映出多数员工培训评估成绩处于中等水平，高分与低分员工均为少数。从晋升角度推测，中等成绩员工基数大但优势不突出；低分员工培训成果欠佳，晋升可能相对小；高分员工培训表现优异，在晋升竞争中或许更具优势。

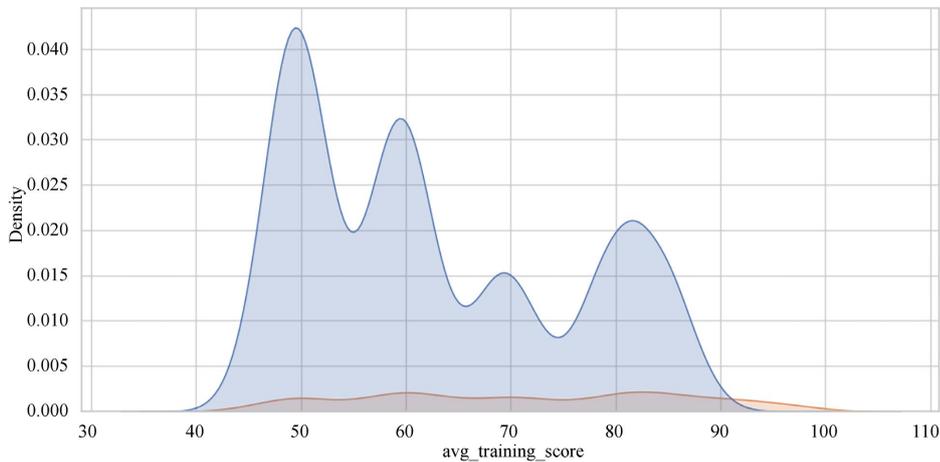


Figure 6. Cross-graph of employee avg_training_score and promotion
图 6. 员工培训评估平均分与晋升情况的交叉图

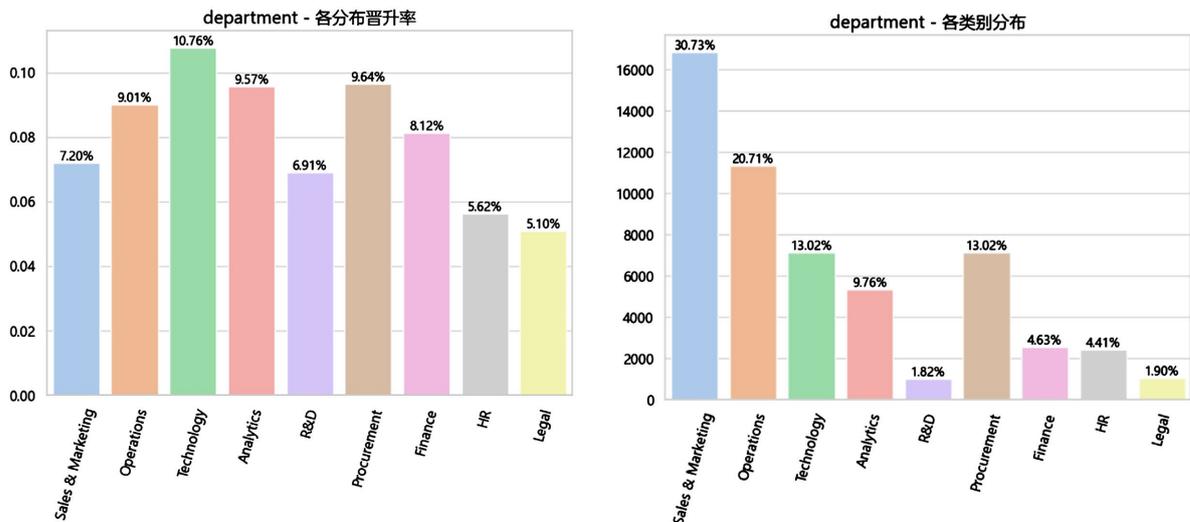


Figure 7. Cross chart of department and promotion status
图 7. 部门与晋升情况的交叉图

(2) 员工所在部门

图 7 是部门与晋升情况的交叉图。图 7 显示：销售和市场部门与运营部门员工数量最多，总共占比超 50%，研发部门员工数量最少，仅占 1.82%。技术部门的晋升率最高，为 10.76%，法务部门晋升率最低，

仅有 5.1%。这可能表明公司在技术领域提供了更多的职业发展机会，以吸引和留住技术人才。人力资源部门和法务部门这些部门的晋升率相对较低，可能是因为这些岗位的流动性较小，或者晋升机会相对有限。

(3) 员工的总绩效表现

图 8 是上一年是否赢得奖项 awards_won?和上一年评级 previous_year_rating 相加后代表总绩效表现的指标 sum_metric 对员工晋升的影响情况。该指标数值越高，代表员工过往绩效表现越优异。分析表明，sum_metric 得分越高，员工的晋升概率呈现显著上升趋势。这说明晋升决策并非孤立看待单一事件(如获奖)或单一评分，而是系统性地评估员工整体的历史绩效表现。优异的综合绩效记录是获得晋升的核心依据，体现了组织在人才选拔中对持续高绩效表现的重视。

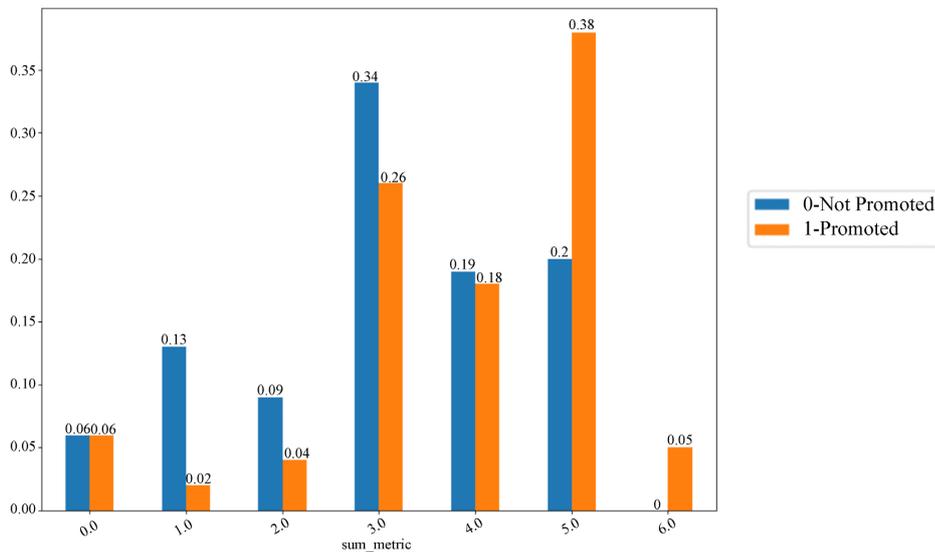


Figure 8. Impact of sum_metric on promotion

图 8. 总绩效表现与晋升情况

(4) 员工的培训次数

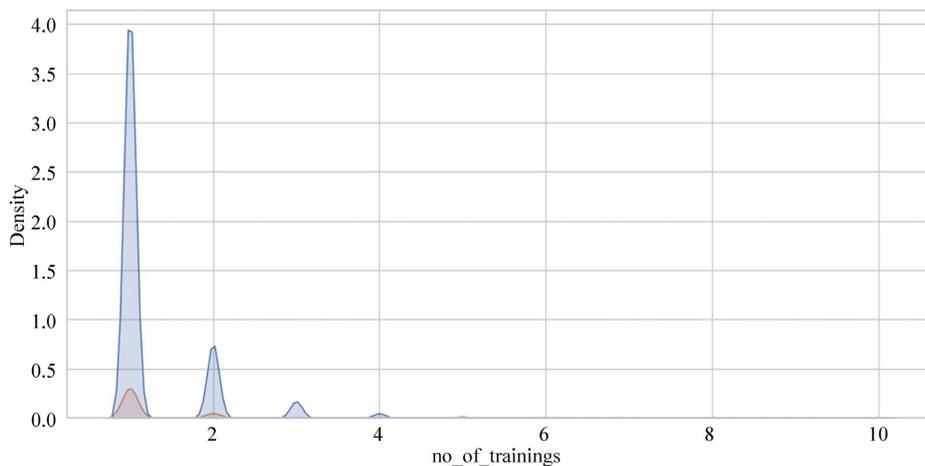


Figure 9. Cross-graph of employee no_of_trainings and promotion

图 9. 员工培训信息与晋升情况的交叉图

图 9 是员工培训次数与晋升情况的交叉图。根据图 9 数据可得：培训次数越少，晋升可能性越高。

原因在于培训的补救性质——公司往往安排绩效未达标的员工参与培训以提升能力。因此，高频次培训可能反映员工持续存在能力短板或绩效问题，间接预示其晋升潜力较低。而培训需求少的员工，通常已具备岗位胜任力或表现出色，更易进入晋升候选池。

(5) 员工的服务年限

图 10 是员工服务年限与晋升情况。根据图 10 发现：拥有五年及以上服务年限的核心员工群体，其晋升情况保持稳定。在入职首年未能实现晋升的新进人员，其未来获得晋升的可能性将显著下降。这一现象说明：第一，组织高度认可并持续投入资源以促进长期服务雇员的职业成长，将晋升作为保留骨干力量、激发其持续贡献的关键策略；第二，新员工入职早期阶段，是甄别高潜力个体并为其铺设加速晋升路径的核心时机。若未能在该阶段展现出突出潜力，后续争取晋升将面临更大阻力。

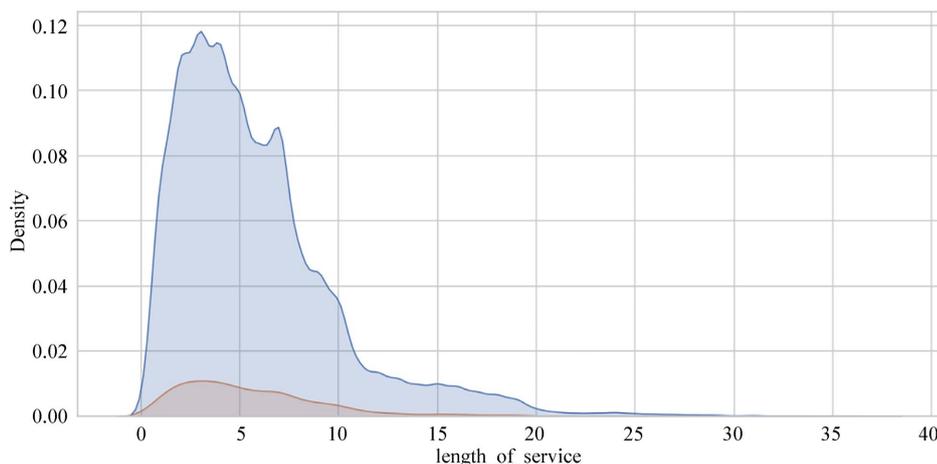


Figure 10. Length_of_service and promotion status

图 10. 员工服务年限与晋升情况

6. 结论

6.1. 研究结论

经过以上研究，本文得出以下结论：

(1) 针对员工晋升预测模型构建，本文对 Kaggle 员工晋升数据集进行了数据清洗、组合、编码、相关性分析和平衡处理，并采用递归特征消除进行降维，最终选定 17 个特征变量。基于此，本文构建并优化了 Logistic 回归、决策树、随机森林、XGBoost 和 Stacking 五种预测模型。在比较了模型各评价指标后，结果显示，集成模型的表现总体优于单一模型，其中 XGBoost 与 Stacking 模型都有良好表现，XGBoost 模型在准确率与精确率上略优于 Stacking 模型，因此本文认为 XGBoost 模型在预测员工晋升问题上性能最好。

(2) 针对模型可解释性与关键影响因素，本文利用 SHAP 值对预测结果进行了可解释性分析，可视化特征影响并排序。结果表明，影响员工晋升的关键前五因素为：培训评估平均得分(avg_training_score)、员工所在部门(department)、总绩效表现(sum_metric)、上一年培训次数(no_of_trainings)和服务年限(length_of_service)。

6.2. 企业优化晋升体系建议

(1) 构建岗位能力匹配型培训评估模型

针对培训评估平均分($avg_training_score$)的强正向影响,建议将培训评估体系与岗位能力矩阵深度绑定:首先,需建立各层级岗位的核心能力清单,如技术岗侧重算法创新能力,管理岗侧重团队协作能力,使培训评估维度(如知识测试、实操模拟)与能力要求一一对应;其次,应引入动态调整机制,每季度根据业务战略更新评估权重,例如数字化转型期增加“数据工具应用”指标权重;此外,还需设置“高潜力员工专项评估通道”,对得分前10%的员工增加跨部门项目实践评分项,以此强化其综合能力信号。

(2) 实施早期职业加速与长期贡献激励计划

鉴于服务年限($length_of_service$)对晋升的阶段性影响,企业需针对不同工龄段员工设计差异化策略:为入职1~3年的员工打造“加速成长通道”,具体包括将轮岗周期压缩至3个月以快速积累多岗位经验,为其配备跨层级导师提供针对性指导;对于服务满5年及以上的核心员工,则推行“贡献积分制”,每满2年即增加5%的晋升评审权重,并且开放“内部创业项目”申请资格,通过将个人晋升与组织长期发展深度绑定,激发其持续贡献动力;而对于入职首年未能实现晋升的员工,应及时启动“潜力重估计划”,在3个月内完成能力诊断并制定个性化改进方案,避免因短期表现误判而错失潜在人才。

(3) 提升晋升决策的科学性与透明度

企业通过开发内部晋升预测系统来提升科学性,根据市场环境和企业战略动态调整评估模型,并利用数据可视化直观展示评估结果。同时,需为有潜力的员工提供领导力培训和轮岗机会积累经验,并鼓励上级基于实际观察进行负责任的推荐。在透明度方面,要公开晋升标准、流程及个人发展计划,公示晋升结果接受监督,并设立申诉复核机制保障公正性。此外,建立反馈机制跟踪新晋升员工表现,收集数据用于持续优化晋升模型,形成改进闭环。

参考文献

- [1] Fernandez-Lozano, I., González, M.J., Jurado-Guerrero, T. and Martínez-Pastor, J. (2020) The Hidden Cost of Flexibility: A Factorial Survey Experiment on Job Promotion. *European Sociological Review*, **36**, 265-283. <https://doi.org/10.1093/esr/jcz059>
- [2] 唐乐, 杨付, 杨伟国. 员工政治技能对晋升机会的影响机制研究[J]. *经济管理*, 2019, 41(10): 73-89.
- [3] 王庆娟, 朱征, 张金成, 等. 晋升公平概念及效用机制的探索——一项质性研究[J]. *南开管理评论*, 2023, 26(5): 157-171.
- [4] 苏屹, 樊起铭, 李丹. 晋升制度模糊化对员工离职倾向的影响研究[J]. *工业工程与管理*, 2018, 23(6): 143-148+172.
- [5] 张敏, 赵宜萱. 机器学习在人力资源管理领域中的应用研究[J]. *中国人力资源开发*, 2022, 39(1): 71-83.
- [6] Raza, A., Munir, K., Almutairi, M., Younas, F. and Fareed, M.M.S. (2022) Predicting Employee Attrition Using Machine Learning Approaches. *Applied Sciences*, **12**, Article 6424. <https://doi.org/10.3390/app12136424>
- [7] 王联英, 慈玉鹏. 基于决策树算法在人力资源推荐技术中的应用[J]. *现代电子技术*, 2021, 44(3): 105-110.
- [8] Cappelli, P., Tambe, P. and Yakubovich, V. (2018) Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3263878>
- [9] Ilwani, M., Nassreddine, G. and Younis, J. (2023) Machine Learning Application on Employee Promotion. *Mesopotamian Journal of Computer Science*, **2023**, 100-114. <https://doi.org/10.58496/mjcs/2023/013>
- [10] Karimi-Majd, A.M., Mahootchi, M. and Zakery, A. (2017) A Reinforcement Learning Methodology for a Human Resource Planning Problem Considering Knowledge-Based Promotion. *Simulation Modelling Practice and Theory*, **79**, 87-99. <https://doi.org/10.1016/j.simpat.2015.07.004>
- [11] Zhang, C., Lv, T., Yuan, C., Ren, Y. and Wang, S. (2020) The Influence of Demographic Characteristics on Employee Promotion: Research Based on Data Mining and Game Theory. *Wireless Communications and Mobile Computing*, **2020**, 1-12. <https://doi.org/10.1155/2020/8814733>
- [12] Jafor, M.A., Wadud, M.A.H., Nur, K. and Rahman, M.M. (2023) Employee Promotion Prediction Using Improved AdaBoost Machine Learning Approach. *AIUB Journal of Science and Engineering*, **22**, 258-266. <https://doi.org/10.53799/ajse.v22i3.781>
- [13] 张辉华. 管理者情绪智力与领导有效性、晋升和离职的关系[J]. *管理科学*, 2012, 25(4): 69-78.

- [14] 周文霞, 谢宝国, 辛迅, 等. 人力资本、社会资本和心理资本影响中国员工职业成功的元分析[J]. 心理学报, 2015, 47(2): 251-263.
- [15] 王建. 工作组织流动、市场能力与职业晋升机遇[J]. 社会学评论, 2021, 9(2): 115-135.
- [16] 张瑞娟, 刘金菊, 尹鹏飞. 什么导致了女性职业晋升障碍?一个基于多理论视角的整合模型[J]. 中国人力资源开发, 2022, 39(11): 90-109.
- [17] 吕芳. 中国女性领导干部的晋升障碍与发展路径——基于对地厅级以上女性领导干部晋升规律的分析[J]. 甘肃社会科学, 2020(6): 220-228.
- [18] Ibarra, H., Carter, N.M. and Silva, C. (2010) Why Men Still Get More Promotions Than Women. *Harvard Business Review*, **88**, 80-85.
- [19] Bode, C., Rogan, M. and Singh, J. (2022) Up to No Good? Gender, Social Impact Work, and Employee Promotions. *Administrative Science Quarterly*, **67**, 82-130. <https://doi.org/10.1177/00018392211020660>
- [20] 马玥. 薪酬差距、晋升概率与业绩晋升敏感性——来自央企企业中层员工的案例证据[J]. 云南财经大学学报, 2021, 37(11): 94-110.