

doi: 10.19920/j.cnki.jmsc.2025.10.005

算法驱动的员工离职预测与干预^①

——基于OW物业公司的研究

朱保全¹, 张志学^{2*}, 万雯³, 陈佳木¹

(1. 万物云空间科技服务股份有限公司, 深圳 518049; 2. 北京大学光华管理学院, 北京 100871;
3. 香港大学经济及工商管理学院, 中国香港 999077)

摘要: 在当前就业形势严峻的背景下, 不少企业仍面临着员工高离职率的问题。如何提早发现具有离职倾向的员工, 了解其诱因, 提前实施干预或表达关怀, 是企业降低员工离职率的重要措施。已有研究虽然探讨了离职倾向员工的表现及其诱因, 但没有得出具有预测能力的定量结论。鉴于此, 本研究基于随机森林算法, 在企业实地场景中, 全面考察员工离职倾向的影响因素, 帮助企业提高预测员工离职倾向的能力。首先, 通过特征工程筛选出重要性较高的特征变量用于机器学习分析; 其次, 通过随机森林算法得出离职因素的重要性排序, 提出预测模型; 同时, 在企业开展准实验对预测模型进行检验; 最后, 通过在企业中的应用进一步验证所提预测模型的有效性。此外, 本研究还检验了离职特征与离职倾向的相关关系, 并发现这些关系与以往研究结论较为吻合。本研究不仅基于随机森林算法全面地研究影响员工离职倾向的因素, 也为企业预先干预、关怀从而降低员工的离职率提供了有益的借鉴。

关键词: 机器学习; 随机森林; 离职倾向; 实地实验; 干预

中图分类号: C931.2; F272.92 **文献标识码:** A **文章编号:** 1007-9807(2025)10-0072-19

0 引言

就业难、失业率高的经济环境下, 不少企业仍面临员工高离职率的难题。员工高离职率会造成企业的直接成本(如招聘和培训新员工)与间接成本(如组织知识和凝聚力文化)^[1], 从而增加企业的经营成本^[2]。因此, 如何降低员工离职率成为企业亟需解决的问题。如果企业提前掌握员工离职倾向诱因, 并预测出未来几个月员工离职可能性, 便能提前采取相应措施进行有针对性地、及时地干预与关怀, 从而减少员工离职可能性, 降低离职率^[3, 4]。

以往组织实践中, 管理人员需亲自掌握下属的状态和动向, 才能了解其离职倾向。由于员工在管理人员面前通常会掩盖自己真实感受, 管理人

员很难及时、准确地关注与干预员工的离职。随着大数据、人工智能的兴起, 智能算法的纯粹理性与高计算能力已使其在许多情况下优于人类的经验^[5]。鉴于此, 学界和业界开始基于公司所存储的数据, 采用相关算法来预测员工离职倾向^[6-8]。例如, Feeley等^[9]利用友谊网络对员工离职率进行了预测; Kao等^[10]运用决策树的方法对员工离职率进行了预测。但相关研究均采用解释性建模, 利用统计模型来验证理论中可观测变量间的关系; 这些研究存在研究样本量小、涵盖影响因素有限、缺乏客观性以及说服力不足等问题。此外, 相关算法研究还存在侧重于模型选择与结果比较, 缺乏对影响特征的识别与排序的关注^[11], 无法达到预警目的等问题。

① 收稿日期: 2022-11-25; 修订日期: 2024-06-10。

基金项目: 国家自然科学基金资助项目(72472002)。

通讯作者: 张志学(1967—), 男, 河南信阳人, 博士, 教授, 博士生导师。Email: zxzhang@gsm.pku.edu.cn

针对上述不足,本研究以国内知名物业企业OW中的物业管家为样本。该公司员工特征、业务、客户评价等信息,会以数据的形式在数据库中沉淀。基于这些数据,本研究采用随机森林算法^②,并结合实践,对该算法进行了优化调整,旨在得出员工离职倾向预测模型^③。这一模型可以让管理人员预知员工的离职倾向,并进行主动关注与干预,不但能降低企业离职率,还能减少管理人员发现问题的难度以及因管理者观察不准确而导致的无效沟通。需要说明的是,虽然预测性模型和解释性模型分别从“预测”和“理解”角度出发,但两者之间并非完全对立^[12]。本研究前期数据预处理、预测性模型设计均基于相关的组织行为与人力资源文献,而非单纯从数据中凭空构建。因此,本研究与以往理论和实证研究密不可分、相辅相成。

本研究的主要贡献如下:理论上,采用随机森林方法处理企业数据池中大量数据,提炼出众多可能影响员工离职倾向的因素。传统员工离职倾向的研究关注在特定场景下“小数据”样本内的拟合能力^[2],难以发现可解释及预测企业中相对复杂现象的系统因素。本研究针对员工离职倾向这一现象进行调研,通过大数据分析,不但提炼出多维度及重要排序,还弥补以往小样本研究解释力不足及样本外的泛化能力较低等问题。同时,本研究规避了传统线性模型的缺陷,提出了较为客观、科学的预测员工离职倾向的方法,从而能够全面分析重要特征变量间的关系。这对传统基于心理测量和计量分析的组织管理的研究提供了必要的补充和新的视角^[13]。实践上,本研究为企业利用数字化进行管理提供了思路与指导。调查发现,众多企业并没有将大数据应用到实际管理中,即

便有了人力资源过程的大数据,企业也不知道如何将其应用到组织管理实践中。本研究融合已有理论与大数据分析技术,选择合理准确的大数据分析方法,精准地开展大数据分析,从而提升工作效率和精准性。这为企业管理实践提供了具体指导。此外,本研究揭示了在员工流失率较高的物业管理企业中,基层服务类员工流失的主要因素以及干预策略,为人力资源部门帮助降低员工流失率提供了具体指导,对其他行业的员工管理具有重要参考价值。

1 文献综述

离职倾向(turnover intention)是指员工的离职意识或想法^[14],是实际工作中自愿离职的一个关键前提^[15]。由于引发员工离职的因素错综复杂,且员工离职会给企业人力、财力带来损失,学术界早已关注员工离职相关问题。例如,Muchinsky和Morrow^[16]通过归纳的方法发现,工作相关因素(组织因素及其导致的工作态度)和个人因素(个性、职业和生活特征等)会导致员工离职。Zeffane^[17]认为制度因素(工作条件、薪资、对组织决策的参与程度和监督等)、个体因素(智力、能力、个人经历、性别、年龄及任期等)和员工对其工作的反应因素(工作满意度、工作投入、工作期望等)会影响员工的离职。Price^[18]系统性地总结与展示了环境、组织、个体三方面因素对员工离职倾向的影响。综合现有文献,本研究将影响员工离职倾向的因素分为个人(如人口统计学信息、工作经验、工作表现)、领导(如领导支持)和组织(如工作压力、薪酬福利)等方面。

1) 人口统计学信息。影响员工离职倾向的人

② 本研究综合考虑、对比了多个模型的性能,包括随机森林、决策树、GBDT、XGBoost和AdaBoost。通过全面的对比分析,本研究发现随机森林在精准率(Precision)、召回率(Recall)、准确率(Accuracy)和AUC(Area Under Curve)等关键指标上都表现优异,故选择随机森林模型。

③ 预测性分析虽然较少出现在现有的组织行为与人力资源的实证研究中,但在学术研究中至关重要。首先,预测性模型没有假设变量之间的关系,而大部分拟合效果好的模型也不是假定函数关系而得来的(如,线性关系、U型关系、指数关系);相对地,解释性模型是对预先所关注的变量提出假设,然后利用数据来进行检验。预测性模型能够发掘出数据中更为复杂的规律,有助于提出并验证数据中存在的规律。其次,预测效果提供了评估解释性模型的新角度,衡量每个变量对预测变量重要程度,提供特征排序,进而检验以往理论的可靠性。

口统计学信息包括年龄、性别、学历、组织内任期、婚姻状况等因素。首先,随着年龄、在组织内工作年限以及工作年限导致的工作嵌入性的增加,员工主动离职倾向将大大减少^[19,20];其次,受教育程度越高的员工越容易离职^[21,22];最后,未婚员工的离职倾向要高于已婚员工^[23]。

2) 工作经验。工作经验是指员工随着时间积累而形成的工作态度、动机以及获得的与工作相关的知识与技能等^[24]。这些相关工作经验与员工的离职倾向呈负相关的关系^[25]。

3) 工作表现。高绩效的员工有较低的离职倾向^[26]。运用社会网络的方法分析员工邮件发送情况表明,离职的经理与周围人的邮件往来较少,且较少参与对话;他们往往会从离职前5个月开始改变沟通行为,增加对外交流、沟通的次数,并且向同行发送“提示”的次数会发生波动^[27]。

4) 薪酬福利。薪酬福利会影响员工的去留。现有研究大多发现薪酬水平与离职倾向呈负相关,但少量研究发现二者呈正相关关系。这类研究的对象通常在企业内任期时间较长,工作经验较为丰富,能力较强且薪资水平较高。其获得进一步培养和加薪的可能性越来越小,容易产生自我创业的动机或谋求更高收入的工作^[28]。同时,员工会将自己的薪酬同他人比较,觉得获得的薪酬越公平,离职的倾向越低^[29,30]。

5) 工作压力。根据压力或紧张视角^[31],以及资源保存视角^[32],工作压力源会引发消极情绪反应,导致工作压力,进而促使员工从工作角色中退出和脱离(即产生离职倾向),以保护自己的心理和情绪资源。

总体来看,以往研究大多通过问卷测评的方式了解员工离职倾向的诱因,缺少大样本以及客观的数据印证与支持。此外,探讨中国企业员工的离职倾向需考虑中国情境下不同于西方情境中的影响因素。因此,在已有文献的基础上,本研究基于中国情境全面且深入地探究哪些因素与员工离职倾向有更为紧密与显著的关联,以及这些因素的重要性排序。

2 数据来源与变量分类

本研究以中国知名物业企业 OW 作为场景。该公司在全国 105 个大中城市为 3 000 多个住宅小区提供高品质的住宅物业服务。在每一个小区,公司建立了以管家为核心的服务体系,实施网格化管理,平均每 500 户业主配置 1 名管家,由管家对客户负责。近年来,公司管家年离职率高于 20%,成为公司人力管理的重要问题。

本研究分析的原始数据集及关键字段示例如表 1 所示。同时,在现有员工离职倾向研究文献与公司情况特征的基础上,将这些数据、字段进行分类,为后续分析和理解提供理论和实践支持。本研究将该公司的管家特征值分为四大维度:个人特征、团队特征、上级特征和项目环境特征(见表 1)。个人变量维度涵盖了人口统计学信息(如性别、年龄、工作年限等)、薪酬福利、工作经验和工作表现。其中薪酬福利包括总收入、提成等相关因素。工作经验用是否为好管家、是否为管家导师、是否有同行经验等字段来表征。这些经验指标代表了员工的知识、技能、能力等信息。工作表现包括排班信息、工单处理信息、物业费收缴与客户评价信息等。团队变量包含团队人数、平均司龄、好管家占比、管家导师占比等。上级变量包含上级司龄、在岗时长、上级绩效表现等。项目环境主要体现在工作压力上,包含员工所服务的项目的交付年限、户数、客户满意度等。由此可见,上述公司后台记录的客观数据,与以往研究所关注的员工离职倾向影响因素相契合。

3 研究方法和模型构建

3.1 特征工程

本研究结合业务场景对数据进行了系统性特征工程处理。通过数据关系的加工和整合,旨在优化数据模型的解释能力和精度,以实现数据价值的最大化。

表 1 数据字段示例
Table 1 Data field example

维度	序号	子维度	数据/数据说明
个人变量	1	基础信息	性别/年龄/司龄/受教育程度等
	2	是否异乡	是否在家乡就职
	3	薪酬福利	员工薪酬信息
	4	工作经验	是否为好管家
	5		是否为管家导师
	6		是否通过大管家资格认证
	7		管家资格认证考试得分
	8		是否有同行经验
	9		是否有转岗/复合经验
	10	工作表现	月度网格满意度(及环比)
	11		月度网格客户投诉量(及环比)
	12		物业费收缴率
	13		管家排班信息
	14		管家你好明细
	15		管家工单明细
	16		业主评价管家信息
团队/组织变量	17	团队规模	该项目管家人数
	18	团队储备	该项目管家储备率
	19	团队能力	该项目好管家占比
	20		该项目导师管家占比
	21		该项目客户对管家线条的满意度
	22	团队稳定性	该项目管家团队平均司龄
	23		近一年本项目管家离职率
上级/领导变量	24	上级资历	大管家司龄
	25	上任时长	大管家在本岗位的任职时长
	26	上级管理	大管家绩效
项目环境条件变量	27	项目规模	户数
	28	项目新旧	交付年限
	29	项目类型	OW 盘/非 OW
	30	项目来源	全委/睿服务/并购/自建
	31	客户关系	项目客户满意度(及同比提升 pcts)
	32		项目月度客户投诉数量及环比
	33	客户期望	物业费单价
	34	期望落差	物业费单价/客户满意度,由高到低

特征工程的目标是从原始数据中最大限度地提取有用特征,以便于算法和模型的应用.数据和特征决定了机器学习的上限,而模型和算法只能逼近这个上限.好的特征一方面能得到较好的效果和更快的运行速度,更容易

理解和维护;另一方面,不需要花太多时间去寻找最优参数,从而降低模型复杂度.特征工程的终极目的是提升模型的性能,包括模型应用效果、执行效率及模型可解释性.图1为特征工程主要内容.

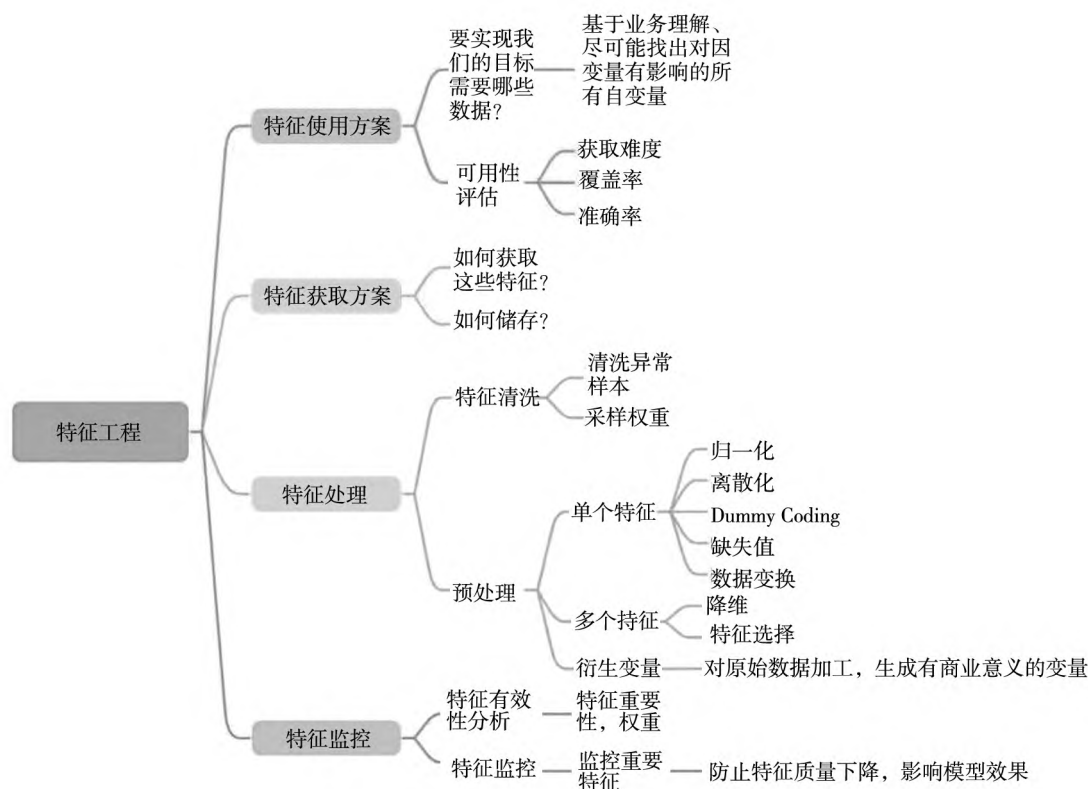


Fig.1 Main contents of the feature engineering

图1展示了系统性特征工程的主要内容,如数据归一化、特征选择、维度压缩等多个方面.本研究选择与离职相关性强的特征(相关性0.25以上),并结合业务专家经验,最终从192个特征中选取了120个重要性高的特征应用于机器学习的部署.输入数据与特征工程示例如图2所示.需要说明的是,特征工程系列工作均基于以往员工离职倾向的相关文献^[13,14],通过相应维度划分,对变量进行归类、整理,而非单纯从数据中凭空构建.具体来说,管家数据分为基本信息和行为数据.其中,基本信息包括管家姓名、年龄等基本信息.行为数据分为管家收缴率、工作表现等方面的数据,还包括管家排班信息、客户评价等相关数据.

3.2 模型训练与选择

管家离职数据是模型中至关重要的因变量,在海量原始数据与特征工程的支持下,研究可以输出特征的重要性排序,即从众多自变量中识别出影响管家离职最为重要的因素;不仅如此,研究通过机器学习,能够实现对每一名管家未来三个月离职概率的预测,并输出基于管家个性化数据的特征模型.当数据的抓取、汇集、处理和输出流程自动化完成后,本研究便进入模型的训练、测试和迭代阶段.为实现更好的预测效果,采用决策树模型及以其为基础构建的集成学习算法对员工离职进行预测.集成学习包括深度神经网络(DNN)、梯度提升决策树(GBDT)、极端梯度提升(XGBoost)和随机森林.

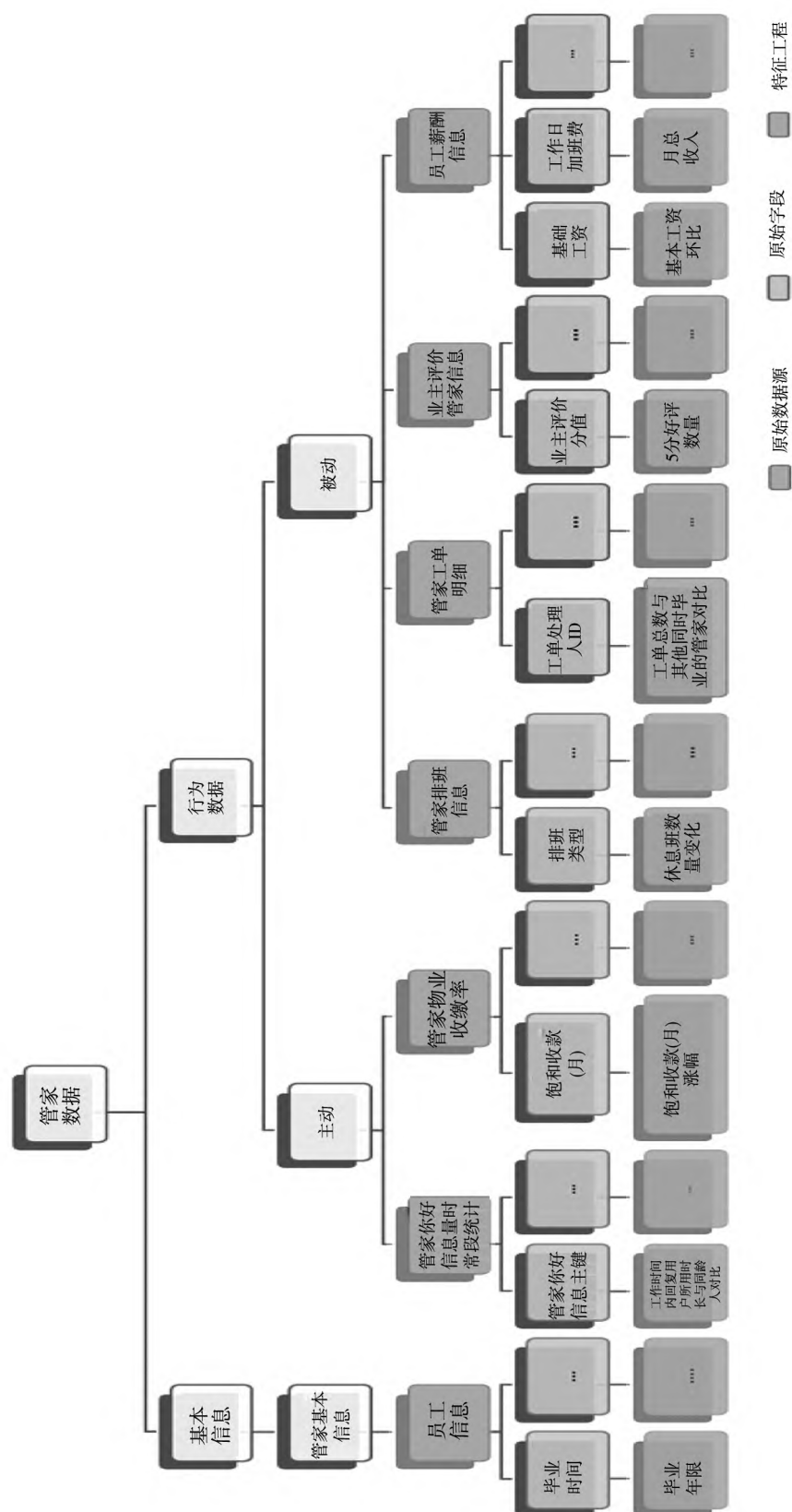


图 2 输入数据及加工示例

Fig. 2 Example of input data and processing

近来, DNN 在各类预测任务中表现出色, 尤其是在处理复杂数据模式和大规模数据集方面. GBDT 是一种通过训练多个决策树并将它们的预测结果进行加权累加来提高性能的集成学习方法. 它准确性高、能有效处理高维数据和缺失值, 但训练时间可能较长, 需要仔细调整参数. XGBoost 是 GBDT 的一个优化版本, 专为速度和性能进行了优化. 它具有非常高的准确性和速度, 可并行化, 并有正则化项防止过拟合. 尽管有默认设置

通常表现良好, 但同样需要仔细调整参数. 随机森林是一个通过训练多个决策树并对其结果进行投票或平均来工作的集成算法. 它准确性高, 能有效处理高维数据, 对缺失值和噪声有良好的容忍度. 经过对比不同训练模型 (DNN、GBDT、XGBoost)^[33], 本研究发现随机森林在精准率 (Precision)、召回率 (Recall)、准确率 (Accuracy) 和 AUC (Area under curve) 等关键指标更优 (见表 2). 因此, 最终选用随机森林这一集成模型.

表 2 模型评估指标对比

Table 2 Comparison of model evaluation indicators

模型	精准率 (Precision)	召回率 (Recall)	准确率 (Accuracy)	AUC
随机森林	90.50%	87.80%	93.60%	0.87
DNN	88.20%	85.90%	91.70%	0.83
GBDT	87.50%	85.60%	91.90%	0.84
XGBoost	89.00%	86.20%	92.10%	0.85

随机森林是一种机器学习算法^[34]. 如果预测结果为离散值, 则为随机森林分类; 如果预测结果为连续值, 则为随机森林回归. 许多研究证实了随机森林算法具有较高的预测精度, 对异常值和噪声具有良好的容忍度. 基于本研究特点, 对随机森林算法进行进一步优化和调整, 具体做法如下:

1) 在特征选择方法上, 结合了基尼指数 (Gini) 和信息增益比 (information gain ratio, IGR), 以提高特征选择的准确性. 信息增益比定义如下:

信息增益 (IG) 为

$$IG(D, A) = H(D) - \sum_{v \in \text{values}(A)} \frac{|D_v|}{|D|} H(D_v)$$

其中 $H(D)$ 为数据集 D 的熵, D_v 为特征 A 取值为 v 的子集.

$$IGR(D, A) = \frac{IG(D, A)}{H(A)}$$

通过结合基尼指数和信息增益比, 本研究提出了一种混合特征选择方法 (hybrid feature selection), 选择特征时考虑两者的加权平均

$$\text{Score}(A) = \alpha \cdot \text{Gini}(A) + (1 - \alpha) \cdot \text{IGR}(D, A)$$

其中 α 为权重参数, 通过交叉验证确定最优值.

同时, 为了提高模型的灵活性, 本研究提出了动态特征子集大小的方法 (dynamic feature subset size), 根据当前节点的数据规模动态调整特征子集的大小. 动态特征子集大小定义为

$$\text{Subset Size} = \lfloor \sqrt{n} \rfloor$$

其中 n 为当前节点的数据样本数.

另外, 为了增强模型的稳健性, 本研究在随机森林中引入了多种类型的基学习器 (Base Learners), 包括决策树、GBDT 和 XGBoost. 通过集成多种基学习器, 提高了模型对不同数据分布的适应能力. 最终预测结果通过加权平均方式得到

$$\text{Final Prediction} = \sum_{i=1}^N w_i \cdot f_i(x)$$

其中 $f_i(x)$ 为第 i 个基学习器的预测结果, w_i 为对应的权重.

2) 超参数调优

为了提高模型的准确性和稳定性, 本研究对随机森林算法进行了超参数调优. 具体优化步骤如下:

① 树的数量 ($n_estimators$): 通过交叉验证确定了最优的树的数量, 从而平衡了模型的偏差和方差.

② 树的深度 (max_depth): 调整树的最大深度, 以防止过拟合, 同时确保模型能够捕捉到数据中的复杂模式.

③ 最小样本分裂 ($min_samples_split$) 和最小样本叶子数 ($min_samples_leaf$): 通过调节这些参数, 控制树的生长, 进一步优化模型的泛化能力.

3) 特征选择

在特征选择过程中,结合了业务专家的经验 和数据分析结果,选取了与离职相关性强的特征 (相关性 0.25 以上),最终从 192 个特征中选取了 120 个重要性高的特征应用于机器学习的部署。

4) 集成学习

为了提升模型的稳定性和预测能力,本研究 尝试了与其他算法集成的学习方法,如结合 GB- DT 和 XGBoost 等方法,通过加权平均的方式,进 一步提升了模型的性能。

5) 模型验证与自动化实现

在模型训练和测试阶段,本研究不仅关注了 模型的预测精度,还注重了模型在企业场景中的

实际应用效果。通过实现数据抓取、汇集、处理、输 出流程的自动化,使模型能够实时更新和优化,并 结合企业实际需求进行动态调整。

此外,本研究还对比小样本学习与大数据在 预测员工离职上的差别、小样本学习和零样本学 习的局限性以及大数据的优势。具体地,本研究大 数据集包含 100 万条员工记录,涵盖多维度特征, 如人口统计信息、工作表现、薪酬福利等;同时从 从大数据集中随机抽取 1 万条记录,保持特征维 度一致,并考虑了在数据稀缺情况下的零样本学 习。结果发现,大样本学习在精准率(*Precision*)、 召回率(*Recall*)、准确率(*Accuracy*)和 *AUC*(*Area under curve*)等关键指标更优(见表 3)。

表 3 大小数据集结果对比

Table 3 Comparison of results between large and small datasets

数据集	模型类型	精准率(<i>Precision</i>)	召回率(<i>Recall</i>)	准确率(<i>Accuracy</i>)	<i>AUC</i>
大数据集	随机森林	90.50%	87.80%	93.60%	0.87
小数据集	随机森林	72.20%	69.50%	74.30%	0.71
小数据集	零样本学习(GAN)	63.10%	61.90%	65.10%	0.67

基于 OW 企业数据池中的数据量,本研究中 机器学习的初始全量数据约 1 亿条记录,其中 70% 用于模型训练,30% 用于模型测试。具体来 说,根据数据的时间戳信息,将整个数据集按时间 顺序划分为训练集和测试集。本研究使用前 70% 的时间段的数据作为训练数据,后 30% 的时间段 的数据作为测试数据。这种划分方式能够模拟实 际预测中未来数据的情况,确保模型在时间维度 上的一致性。为了确保训练数据和测试数据在人口 特征学变量上没有显著性差异,通过独立样本 *t* 检验和卡方检验发现,训练数据和测试数据在人口 特征学变量上没有显著性差异。研究中在模型 训练阶段对各方法均进行充分调优,测试其在测 试集上的预测效果。

4 管家离职率预测的实验与效果 评估

4.1 预测模型评估

模型的评估是以预测情况与真实情况之间的 拟合度为依据的。所有被预测的样本都可以归为 以下四类:预测离职且实际离职,命名为 True Positive(真正, *TP*);预测离职但实际未离职,命名 为 False Negative(假负, *FN*);预测不离职且实际 不离职,命名为 True Negative(真负, *TN*);预测不 离职但实际离职,命名为 False Positive(假正, *FP*),如表 4 所示:

表 4 模型评估的基础框架

Table 4 Basic framework for model evaluation

变量		预测情况	
		离职(需要关怀)	不离职(不需要关怀)
真实情况	离职(需要关怀)	True Positive(<i>TP</i>)	False Negative(<i>FN</i>)
	不离职(不需要关怀)	False Positive(<i>FP</i>)	True Negative(<i>TN</i>)

表 5 预测评估模型人数

Table 5 Predictive model for evaluating population numbers

变量	预测: 离职(需要关怀)	预测: 不离职(不需要关怀)	总人数
实际: 离职(需要关怀)	181	499	680
实际: 不离职(不需要关怀)	127	6 671	6 798
总人数	308	7 170	7 478

本研究中,模型评估的关键指标得分、相应人数如表 5 所示

1) 精准率(*Precision*). 代表预测结果的精准度,指模型预测为“需要关怀”的人中,实际也需要关怀的占比. $Precision = TP / (TP + FP)$. 例如, 2019 年 1 月份时,模型预测未来三个月,即 2019 年 1 月—2019 年 3 月有 308 人需要关怀. 而实际上,到了 2019 年 3 月份结束时,发现这 308 位员工中,有 181 人是真正需要关怀的,即 *Precision* 是 $181 / 308 * 100\% = 58.8\%$.

2) 召回率(*Recall*). 即覆盖面的度量,也称为击中概率或查全率,表示实际需要关怀的人中,被模型预测到需要关怀的比例. $Recall = TP / (TP + FN)$. 例如, 2019 年 1 月—2019 年 3 月有 680 人实际上需要关怀,而 2019 年 1 月份模型预测未来三个月需关怀的人员中,包含了这 680 人中的 181 人,即 *Recall* 是 $181 / 680 * 100\% = 26.6\%$.

表 6 预测评估模型关键指标得分

Table 6 Key indicator score of predictive evaluation model

指标	计算公式	百分比
精准率(<i>Precision</i>)	$TP / (TP + FP)$	58.80%
召回率(<i>Recall</i>)	$TP / (TP + FN)$	26.60%
准确率(<i>Accuracy</i>)	$(TP + TN) / (P + N)$	91.60%
特异性(<i>Specificity</i>)	$TN / (TN + FP)$	98.10%

最终,本研究选择用 *AUC* 指数作为模型估计的判定值. *AUC* 是机器学习领域中评估模型的指标,指 ROC 曲线与 FP Rate 轴围成的面积(横轴为 FPR,纵轴为 TPR). 在机器学习中,没有理想的或者可接受的 *AUC* 分数,不同的问题场景对 *AUC* 效果的要求不尽相同,甚至全然不同. *AUC* 越大,说明性能越好;当 $AUC < 0.5$ 模型错误;当 $AUC = 0.5$ 则根据现实场景来判断. 本研究中预测模型的 *AUC* 值为 0.8,证明该预测模型良好,可在后续管理实践中使用.

4.2 预测模型的阈值设定

根据算法输出每一名管家在未来三个月的离

3) 准确率(*Accuracy*). 即被分对的样本数占所有样本数的比例,表示模型正确地预测了人员去向的比例. $Accuracy = (TP + TN) / (P + N)$. 例如 2019 年 1 月份时,模型预测未来三个月有 308 人需要关怀,7 170 人不需要关怀;实际上 2019 年 1 月—2019 年 3 月,被预测需要关怀的 308 人中,有 181 人真正需要关怀;被预测不需要关怀的 7 170 人中,有 6 671 人实际上也不需要关怀. 即模型预测的准确率是 $(181 + 6 671) / (308 + 7 170) * 100\% = 91.6\%$.

4) 特异性(*Specificity*). 表示实际上不需要关怀的员工中,被模型预测为“不需要关怀”的占比;与特异性相对的指标为虚惊概率(False-Positive Rate),指实际不需要关怀的员工被模型预测为“需要关怀”的占比,二者的关系为: $Specificity = TN / (TN + FP) = 1 - FPR = 1 - 127 / 6 798 = 98.1\%$ (见表 6).

职概率,OW 物业公司将对离职概率超过一定阈值的管家实施关怀干预. “关怀阈值”是本研究设定的一个标准,用于判断员工是否需要接受特定的关怀行为以降低其离职风险. 这个阈值是基于本研究的预测模型计算出的员工离职风险得分. 当员工的离职风险得分超过这个阈值时,就需要对其实施关怀行为.

关怀阈值的确定,对公司实施干预的成本及对模型效果的评估都至关重要. 如果关怀阈值设置过高,模型预测失误的风险就会大大降低(准确率越高),被预测为“离职”的管家数量,即组织关怀所覆盖的群体数量就会越少(覆盖率越低),

模型的实际价值便会大打折扣;如果关怀阈值设置过低,就会有更多的管家被预测为“离职”,被组织实施关怀的管家数量增多,但会由于预测失误差率增加而造成组织资源浪费。

因此,关怀阈值的确定需要综合考虑模型的覆盖率与准确率,以使得模型效果达到最优。本研究模型基于公司实际情况,将最佳阈值设定为 0.5^⑤,即当管家未来三个月离职概率的预测值为 0.5 及以上水平时,组织将启动关怀干预,以降低管家的实际离职数量。当阈值设定为 0.5 时,模型的精准率(Precision)和召回率(Recall)达到较为平衡的状态,即在保证预测准确性的同时,覆盖了

更多需要关怀的员工(见表 7)。

表 7 不同阈值下模型的精准率和召回率

Table 7 Precision and recall of the model at different thresholds

阈值	精准率(Precision)	召回率(Recall)
0.3	45.70%	35.20%
0.4	52.30%	30.40%
0.5	58.80%	26.60%
0.6	65.20%	22.80%
0.7	70.30%	19.40%

4.3 预测模型准实验验证

为了进一步验证该离职倾向预警模型的可信度,本研究还进行了 AB-Test 检验,即通过准实验的方式,对所选模型进行验证。具体操作如下:

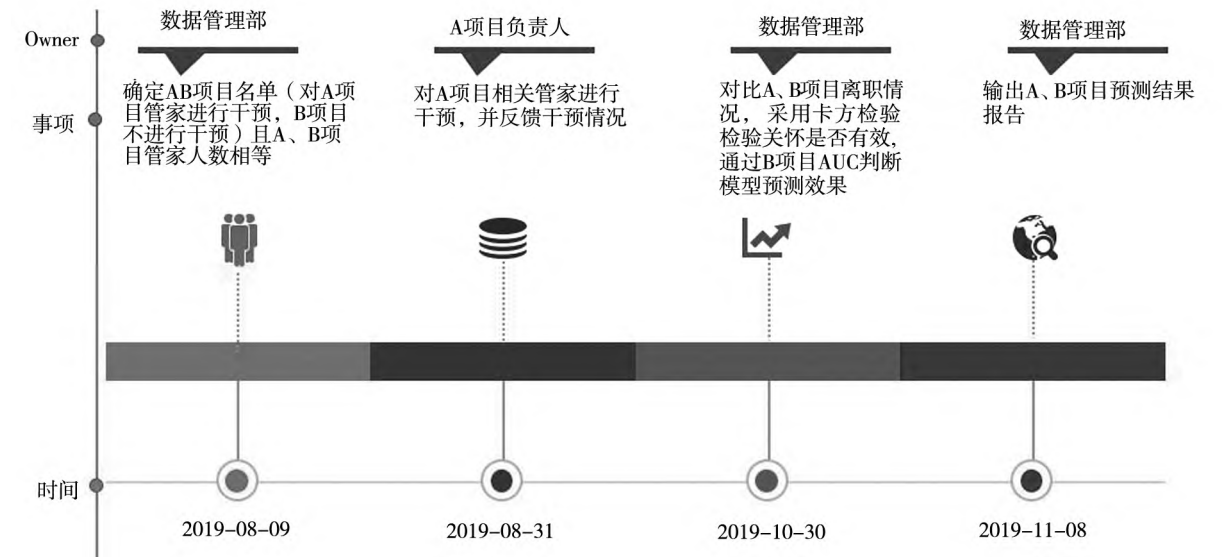


图 3 AB 实验步骤

Fig. 3 Steps of the AB-Test

1) 首先预测出 A、B 项目可能离职的管家,根据实际离职情况将预测离职的管家进行分类。本次实验将全国 14 个试点项目,随机分为 A、B 两组,对 A 组进行干预,对 B 组不进行干预,如表 8

所示。

2) 假设检验

假设 H_0 关怀对管家离职无影响

假设 H_1 关怀对管家离职有影响

表 8 AB—测试

Table 8 AB-Test

组别	实际离职人数	实际未离职人数	合计
A 项目: 干预(进行关怀)	5(n_{11})	114(n_{12})	119($n_{11} + n_{12}$)
B 项目: 未干预(不进行关怀)	15(n_{21})	106(n_{22})	121($n_{21} + n_{22}$)
合计	20($n_{11} + n_{21}$)	220($n_{12} + n_{22}$)	240($n_{11} + n_{12} + n_{21} + n_{22}$)

⑤ 在二分类问题中 0.5 常常被选择作为默认阈值。在这个阈值上,假设正例和负例发生的可能性是相等的。同时,与其他阈值相比,阈值 0.5 时,本研究模型在训练集和测试集上效果最好。

3) 计算自由度与理论频数

自由度 $df = (r-1)(c-1)$,在此场景中 $r = 2$, $c = 2$;

理论频数 $e_{ij} = F_{Y_i} \times F_{X_j} / n$

4) 计算卡方统计量

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}} \sim \chi^2(df)$$

5) 确定 p 值并分析

查找 χ^2 分布临界值表 ,找到卡方值对应的 p 值. 若 p 值 $\leq \alpha$,则拒绝 H_0 ,认为关怀对管家离职有影响; 若 p 值 $> \alpha$,则接受 H_0 ,认为关怀对管家离职无影响. 假设显著性水平 α 一般取 0.05 . 在本研究中 p 值为 0.022 ,小于 0.05 ,故拒绝原假设 ,即关怀对管家离职有影响.

6) 判断影响方向

既然已经判断关怀对管家离职有影响 ,然后比较 A 项目中 $\frac{n_{11}}{n_{11} + n_{12}}$ 的比值与 B 项目中 $\frac{n_{21}}{n_{21} + n_{22}}$ 的比值. 若 $\frac{n_{11}}{n_{11} + n_{12}} < \frac{n_{21}}{n_{21} + n_{22}}$,则说明关怀行为导致离职率降低; 若 $\frac{n_{11}}{n_{11} + n_{12}} > \frac{n_{21}}{n_{21} + n_{22}}$,则说明关怀行为导致离职率升高. 本研究结果发现 A 组的实际离职/预测离职的比率小于 B 组的比率 ,说明关怀行为导致离职率的降低. 因此 ,通过准实验的方式 ,验证了本研究的预测模型是可信的.

该公司管家离职预测模型于 2019 年 6 月正式上线投入使用. 通过一个月的适应学习 ,并熟悉相关关怀内容与方法 ,从 2019 年 8 月开始 ,管家月离职率呈逐月减少趋势. 截至 2019 年底 ,月离职率从 7 月份的 5% 降低到 2% ,离职率减少 3% (见图 4) .

4.4 管家离职因素的重要性排序

本研究基于随机森林进行的模型训练 ,通过

分类与整合 ,得到了 10 个变量的重要性排序 ,旨在为后续管家离职关怀提供有针对性的指导 ,如表 9 所示^⑥. 从最终提取出的重要性特征可以看出 ,影响员工离职的因素主要包含与薪酬相关特征、人口统计学变量、工作压力、绩效表现四个类型.

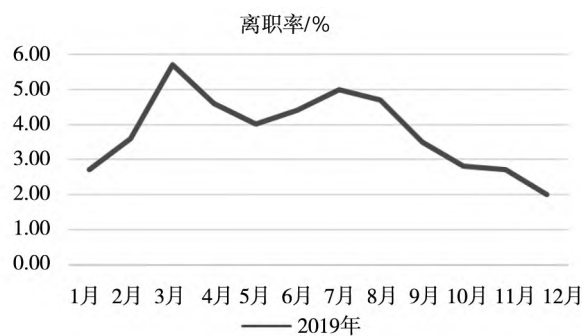


图 4 离职率预测模型应用结果

Fig. 4 Application results of the turnover rate predictive model

表 9 影响管家离职的特征重要性排序

Table 9 Ranking of feature importance affecting property coordinator

排序	特征变量	得分
1	综合收入排名	0.08
2	月度总收入	0.07
3	用户评分排名	0.05
4	工作年限	0.04
5	工作绩效	0.03
6	回应时间	0.02
7	管理中心	0.02
8	年龄	0.02
9	性别	0.02
10	绩效环比	0.02

4.5 管家离职率预测模型应用

基于上述结果 ,OW 物业公司将模型预测结果输出为可视化的管家离职预警平台 ,在这个平

⑥ 在随机森林中 ,特征重要性是通过计算每个特征在所有树中减少不纯度的平均值来得到的 ,这种减少的不纯度通常通过基尼不纯度或信息增益来衡量. 以下是计算特征重要性的基本公式: 特征重要性 = $\sum (\text{节点不纯度} - \text{左子节点不纯度} - \text{右子节点不纯度}) / \text{总节点数}$. 这里 ,节点不纯度可以是基尼不纯度或信息增益 ,左子节点和右子节点是通过特征分裂得到的两个子节点.

台上,模型按月向一线组织输出预警管家离职人数、预警管家离职率(见图5),图中展示的是2020年5月的数字,流失预警人数为490人,流失预警比例为5.9%,及离职倾向高的管家名单(见图6),供管家专家提前实施员工关怀;根据预测离职的特征模型,管家专家还能查看预警离职管家的“关怀要点提醒”(见图7);预警平

台的应用,使得一线管理者对员工离职的管理从被动变为主动,经过模型识别出的离职预警名单,还提高了一线管理者实施员工关怀的效率。例如,当没有模型时,一线管理者识别出52个离职人员,共需要访谈569人;而在模型的帮助下,只需要关怀到327人,管理效率提升了42.5%(见图8)。

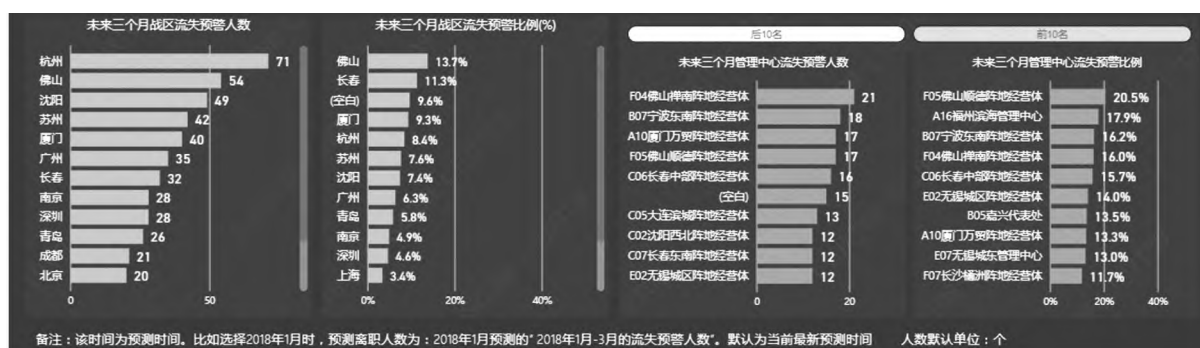


图5 预警管家离职人数

Fig.5 Early warning of the number of property coordinators' turnover



图6 离职倾向高的管家名单

Fig.6 List of property coordinator with a high tendency of turnover

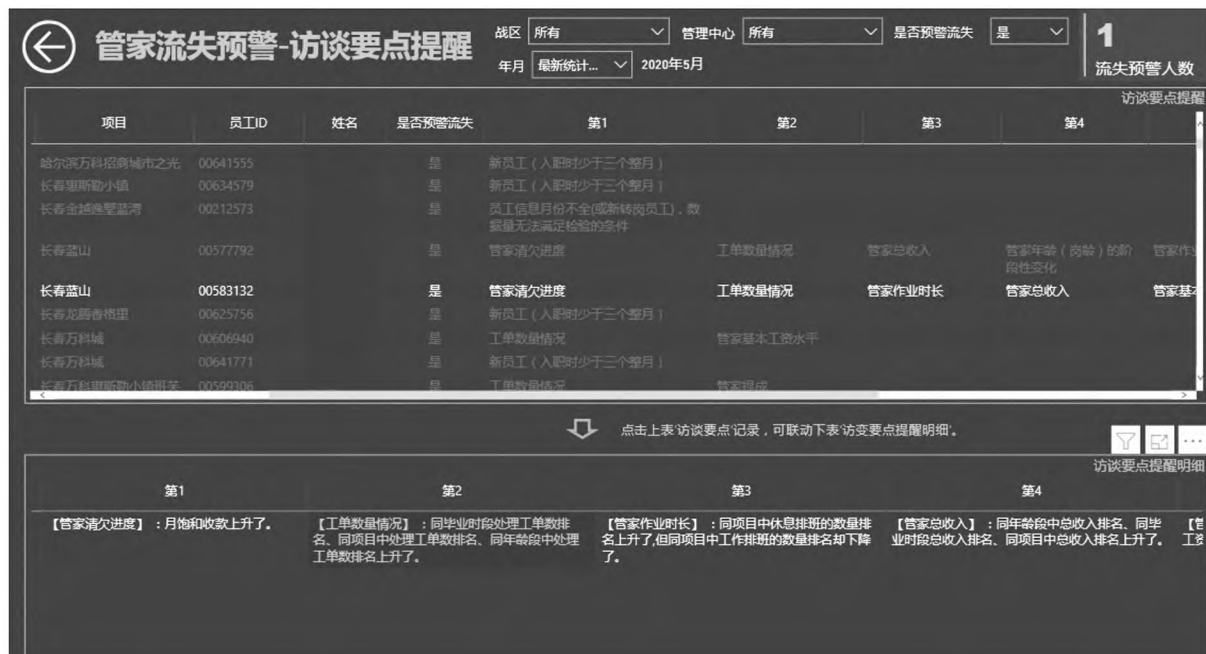


图 7 关怀要点提醒

Fig. 7 Care point reminder



图 8 有无模型管理效率对比

Fig. 8 Comparison of model management efficiency with or without it

对于关怀的方法, 本研究将该公司收集的离职原因, 进行了整理和归纳, 最终发现离职干预应包括以下四种方式: 专家面谈、关心沟通; 安排专人进行辅导; 关注员工动态、给予关注和鼓励; 调整工作内容或工作量。同时, 为了验证归类是否正

确且全面, 从 14 个试点项目中选择了 96 名管家专家, 对离职干预措施进行了进一步调查, 将之前概括的四种方式作为必选措施, 同时进行开放性填写, 作为自选措施。14 个试点项目干预措施情况如表 10 所示。

表 10 14 个项目试点管家离职干预措施

Table 10 Intervention measures for the property coordinators' turnover in 14 pilot projects

措施	数量	百分比
专家面谈、关心沟通	22	50.00%
安排专人进行工作辅导、培训或协助	7	15.91%
关注员工动态 给予关注和鼓励	4	9.09%
调整工作内容或工作量	3	6.82%
给予专项挑战性工作 ,让管家发挥最大价值	2	4.55%
梳理网格疑难事项 ,重点客户分层级 ,分别由项目驻场和管家经理维护 ,减轻网格管家工作压力	1	2.27%
调整薪资	1	2.27%
培养员工自信	1	2.27%
同意离职	3	6.82%
总计	44	100.00%
无、否、0 或空值(必选措施列中)	69	
无、否、0 或空值(自选措施列中)	87	

同时 ,为了更准确且全面地掌握全国项目干预措施的内容 ,本研究对全国项目中的 773 位管家专家进行了离职干预措施的调查. 调查形式与 14 个点项目一致 ,发现全国管家离职干预措施如表 11 所示. 基于调查形成的干预措施编制成管家专家关怀手册 ,旨在为专家进行离职关怀提供指导与借鉴.

表 11 全国管家项目离职干预措施

Table 11 Intervention measures of property coordinator project across the country

措施	数量	百分比
关注员工动态 给予关注和鼓励	123	46.07%
专家面谈、关心沟通	55	20.60%
安排专人进行工作辅导、培训或协助	64	23.97%
调整工作内容、工作量	6	2.25%
更换项目 ,岗位不变	3	1.12%
调整薪资	3	1.12%
给予专项挑战性工作 ,让管家发挥最大价值	2	0.75%
调整工作方式	1	0.37%
增加团建活动 ,释放工作压力	1	0.37%
给予三保业务支持 ,资源倾斜	1	0.37%
公司文化宣贯	1	0.37%
培养管家自信	1	0.37%
对项目遗留问题分类汇总 ,设计统一说辞 ,减轻管家对客户沟通压力	1	0.37%
定期召开项目遗留问题推进会议 ,让管家了解工作进展情况 ,以便与客户沟通	1	0.37%
梳理网格疑难事项 ,重点客户分层级 ,分别由项目驻场和管家经理维护 ,减轻网格管家工作压力	1	0.37%
同意离职	3	1.12%
总计	267	100.00%
无、否、0 或空值(必选措施列中)	521	
无、否、0 或空值(自选措施列中)	711	

5 特征值相关性检验

本研究还利用机器学习进行相关分析,旨在验证通过大数据分析得出的结论是否与以往实证研究结论一致,分析异同点;同时希望通过具体分析出重要特征变量对离职倾向影响的正负相关性,给予管理人员更加准确地指导(见表12)。

表 12 相关性分析

Table 12 Correlation analysis

序号	特征变量	相关性
1	月度总收入	-0.08**
2	被用户评5分数量在工作省份排名	-0.03*
3	管家24小时内回复信息的数量	-0.09**
4	基本工资	-0.07*
5	已工作多久	-0.08**
6	排班类型为工作的数据环比	-0.05*
7	被用户评2分的数量在本管理中心的排名	-0.05*
8	排班时间为休息的总数在同年龄段的排序	-0.05*
9	性别	-0.11**
10	年龄	-0.20***

注: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ 。

从人口统计学相关特征来看,毕业和工作年限越长以及年龄越大的管家离职倾向越低。这一结果与以往实证研究发现基本保持一致。例如, Somers^[21], Cotton 和 Tuttle^[22] 等研究表明,随着年龄和在组织内工作时间的增加,员工的离职倾向会大幅度降低。本研究还发现男性员工的离职率要高于女性。这是因为物业管理对形象思维和表达能力要求较高,且要求认真细致,所以男性员工离职倾向较高^[35]。值得关注的是,本研究的预测模型中并没有包含学历程度这一人口统计学变量。但现有研究大多认为,受教育程度越高的员工,越容易离职^[19]。出现这种差异的原因可能是本研究测试的目标群体是物业管家,大都学历程度不高且学历层次较为均衡。因此,本研究未发现学历所导致的差异化影响。

从薪酬角度来看,基本工资越高、月薪越高的管家离职倾向越低。这一结果与以往研究相一致。一项元分析研究表明,薪酬水平和离职倾向呈负相关关系,相关系数在 $-0.027 \sim -0.678$ 之

间^[36]。因此,对于物业管家来说,薪酬的高低与其离职倾向呈负相关关系。

从绩效与工作表现方面来看,管家的工作班次、用户点评分排名等越低,表明其离职倾向越高。通过这些特征指标,可以看出员工工作的投入以及努力、努力程度,与其离职倾向负相关。因此,在绩效方面,本研究预测结果与现有研究一致^[37]。

6 结束语

如何降低员工离职率已成为企业人力部门的工作之重,也是学界长久以来所重点关注的话题。本研究以中国知名物业企业 OW 中物业管家这一群体为样本,通过机器学习的方法揭示影响员工离职的重要因素,并实施精准干预。首先,通过该公司数据池中的海量数据,进行特征工程筛选,并运用随机森林的算法对员工离职倾向的影响模型进行预测,得到 10 个对管家离职倾向有重要的重要特征变量及其重要性排序。其次,对预测模型进行评估,通过实地实验对模型进行验证。最后,将被验证过的模型在全公司范围内应用,针对有离职倾向的员工开展干预。本研究在分析现有员工离职倾向相关研究的基础上,通过机器学习中的随机森林的算法,全面考察了员工离职倾向的相关因素,评估了不同因素对员工离职倾向的预测能力,丰富了员工离职倾向这一领域的文献。此外,本研究采用随机森林算法规避了传统线性模型的缺陷。

本研究展现了机器学习在管理研究中的应用,丰富且验证了数智技术在人力资源管理实践中积极作用的呼吁^[38,39]。与纯粹通过数据驱动的机器学习不同,一方面,本研究在梳理已有关于员工离职的文献基础上,结合公司记录的特征数据,将理论和大数据处理技术结合起来,不但获得了预测准确性,还进一步证实了以往研究发现的有效性,以及小样本学习和零样本学习的局限性以及大数据的优势。另一方面,运用随机森林的算法对员工离职倾向的影响模型进行预测,并通过在

企业从事的现场实验对预测模型进行评估。这对传统基于心理测量和计量分析的组织领域研究提供必要的补充和新的视角^[13]。此外,本研究代表了智能时代企业管理研究的一个方向和路径。通过机器学习处理海量的数据,并基于已有的研究发现,本研究对特定管理现象的多维度、跨领域的大数据进行分析,能够更准确地把握与现象最有关联的因素,从而建立更具解释力的理论。

在实践上,本研究探究了员工离职倾向影响因素的重要程度,分析了相对重要变量对员工离职倾向的作用效果。通过机器学习获得的特征重要性排序向物业公司传递了重要的工作启示。由于传统企业是由管理者履行管理职能,而他们受到动机驱使或囿于认知偏差很容易出现判断失误,导致员工离职倾向预测的精准度不高。随着企业智能系统的完善及大数据采集、存储技术的进步,传统企业人力资源领域的多种数据日益可测可获,能为大数据分析提供丰富的原材料,进而借助机器学习、人工智能等方法解决员工离职倾向预测率较低的问题。运用智能技术对于大数据分析后获得以下重要发现:首先,薪酬收入对于管家群体是非常直接与重要的激励因子,关注管家收入现状,赋能管家创造更多收入总和,让管家感知个人薪酬收入的进步,是重要的管理抓手;与之相应地,较高的物业费收缴率不仅是较好工作能力、较好客户关系的表现,也能较为明显地降低管家

工作压力和心理负担,从而促进管家保有;其次,物业管具有高度社区嵌入性,工作年限越长的管家能更深地嵌入在社区中^[20],从而具有较低的离职倾向。最后,公司应在工作排班波动时,给予管家更多的关怀与引导,让管家在工作强度增加的同时能够感受到组织的关怀与支持,并且注意劳逸结合,合理安排休息时间。这些相关性关系的发现在实证以往研究结论的同时,也为企业相应管理人员进行实际关怀与干预提供了一定的指导。

本研究还存在不足:首先,就研究数据而言,由于公司数据积累和记录的问题,现有特征指标分析时还缺乏一些管家对工作环境满意度、管家就职情况、培训次数等相关信息。同时,也缺少管家直属领导对其帮助、支持,以及企业外部情境因素(如再就业机会、行业内供求情况等)等影响方面的数据。未来的研究可以在建立客观指标池的同时,关注管家的主观感受以及企业外部因素,进而更为全面的管理离职倾向进行预测,更全面地推进人机融合^[40];其次,在模型建立方面,本研究通过相关算法比较,应用了随机森林算法。随着算法不断地改进,未来研究可以比较不同算法预测的准确率,不断推进机器学习在企业实践中的应用^[39,41];最后,本研究仅关注了物业管家这一服务群体的离职倾向,该预测模型是否适用于其他企业以及其他类型从业人员还有待进一步验证。

参考文献:

- [1] Robinson R N S, Kralj A, Solnet D J, et al. Thinking job embeddedness not turnover: Towards a better understanding of frontline hotel worker retention[J]. *International Journal of Hospitality Management*, 2014, 36: 101 – 109.
- [2] DiPietro R B, Moreo A, Cain L. Well-being, affective commitment and job satisfaction: Influences on turnover intentions in casual dining employees[J]. *Journal of Hospitality Marketing & Management*, 2020, 29(2): 139 – 163.
- [3] Gabrani G, Kwatra A. Machine learning based predictive model for risk assessment of employee attrition[C]. *International Conference on Computational Science and Its Applications*. Cham: Springer, 2018: 189 – 201.
- [4] Li G G, Long J D, Wang F F. A study on turnover intention of knowledge staffs based on the psychological contract[C]. 2013 10th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICC-WAMTIP). New York: IEEE, 2013: 298 – 301.
- [5] 陈冬梅, 王俐珍, 陈安霓. 数字化与战略管理理论——回顾, 挑战与展望[J]. *管理世界*, 2020, 36(5): 220 – 236.
Chen Dongmei, Wang Lizhen, Chen Anni. Digitalization and strategic management theory: Review, challenges and prospects[J]. *Journal of Management World*, 2020, 36(5): 220 – 236. (in Chinese)

- [6] Artelt A , Gregoriades A. “How to make them stay?” Diverse counterfactual explanations of employee attrition [C]. In Proceedings of the International Conference on Enterprise Information Systems , Prague: ICEIS ,2003: 24 –26.
- [7] Marín Díaz G , Galón Hernández J J , Galdón Salvador J L. Analyzing employee attrition using explainable AI for strategic HR decision-making [J]. Mathematics ,2023 ,11(22) : 4677 –4702.
- [8] Marsh E , Vallejos E P , Spence A. The digital workplace and its dark side: An integrative review [J]. Computers in Human Behavior ,2022 ,128: 107118.
- [9] Feeley T H , Hwang J , Barnett G A. Predicting employee turnover from friendship networks [J]. Journal of Applied Communication Research ,2008 ,36(1) : 56 –73.
- [10] Kao H W , Lin S W , Wan S Y. Applying decision tree to predict nursing turnover: A case study in a public hospital [J]. The Journal of Taiwan Association for Medical Informatics ,2012 ,21(4) : 15 –29.
- [11] Gao X , Wen J , Zhang C. An improved random forest algorithm for predicting employee turnover [J]. Mathematical Problems in Engineering ,2019 , (1) : 4140707.
- [12] 陆 瑶 , 张叶青 , 黎 波 , 等. 高管个人特征与公司业绩——基于机器学习的经验证据 [J]. 管理科学学报 ,2020 ,23(2) : 120 –139.
- Lu Yao , Zhang Yeqing , Li Bo , et al. Managerial individual characteristics and corporate performance: Evidence from a machine learning approach [J]. Journal of Management Sciences in China ,2020 ,23(2) : 120 –139. (in Chinese)
- [13] Yarkoni T , Westfall J. Choosing prediction over explanation in psychology: Lessons from machine learning [J]. Perspectives on Psychological Science ,2017 ,12(6) : 1100 –1122.
- [14] Akgunduz Y , Eryilmaz G. Does turnover intention mediate the effects of job insecurity and co-worker support on social loafing? [J]. International Journal of Hospitality Management ,2018 ,68: 41 –49.
- [15] Steel R P , Ovalle N K. A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover [J]. Journal of Applied Psychology ,1984 ,69(4) : 673 –686.
- [16] Muchinsky P M , Morrow P C. A multidisciplinary model of voluntary employee turnover [J]. Journal of Vocational Behavior ,1980 ,17(3) : 263 –290.
- [17] Zeffane R. Patterns of organizational commitment and perceived management style: A comparison of public and private sector employees [J]. Human Relations ,1994 ,47(8) : 977 –1010.
- [18] Price J L. Reflections on the determinants of voluntary turnover [J]. International Journal of Manpower ,2001 ,22(7) : 600 –624.
- [19] Raza A , Munir K , Almutairi M , et al. Predicting employee attrition using machine learning approaches [J]. Applied Sciences ,2022 ,12(13) : 6424 –6441.
- [20] Lee T W , Burch T C , Mitchell T R. The story of why we stay: A review of job embeddedness [J]. Annual Review of Organizational Psychology and Organizational Behavior ,2014 ,1(1) : 199 –216.
- [21] Somers M J. Modelling employee withdrawal behaviour over time: A study of turnover using survival analysis [J]. Journal of Occupational and Organizational Psychology ,1996 ,69(4) : 315 –326.
- [22] Cotton J L , Tuttle J M. Employee turnover: A meta-analysis and review with implications for research [J]. Academy of management Review ,1986 ,11(1) : 55 –70.
- [23] 陈忠卫 , 张广琦 , 秦慈进. 婚姻影响知识型员工的离职倾向吗? ——基于可雇佣性视角的案例研究 [J]. 经济管理 ,2014 , (2) : 179 –188.
- Chen Zhongwei , Zhang Guangqi , Qin Cijin. Does marriage affect the turnover intention of knowledge workers? Case study from the perspective of employability [J]. Business and Management Journal ,2014 , (2) : 179 –188. (in Chinese)
- [24] Tesluk P E , Jacobs R R. Toward an integrated model of work experience [J]. Personnel Psychology ,1998 ,51(2) : 321 –355.
- [25] Sajjadi S , Sojourner A J , Kammeyer-Mueller J D , et al. Using machine learning to translate applicant work history into

- predictors of performance and turnover[J]. *Journal of Applied Psychology*, 2019, 104(10): 1207–1225.
- [26] Griffeth R W, Hom P W, Gaertner S. A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium[J]. *Journal of Management*, 2000, 26(3): 463–488.
- [27] Gloor P A, Colladon A F, Grippa F, et al. Forecasting managerial turnover through e-mail based social network analysis[J]. *Computers in Human Behavior*, 2017, 71: 343–352.
- [28] 张建琦, 汪凡. 民营企业职业经理人流失原因的实证研究——对广东民营企业职业经理人离职倾向的检验分析[J]. *管理世界*, 2003, (9): 129–135.
- Zhang Jianqi, Wang Fan. Empirical study on the reasons for the loss of professional managers in private enterprises: An examination and analysis of the resignation tendency of professional managers in private enterprises in Guangdong[J]. *Journal of Management World*, 2003, (9): 129–135. (in Chinese)
- [29] Martin C. Explaining labour turnover: Empirical evidence from UK establishments[J]. *Labour*, 2003, 17(3): 391–412.
- [30] Alterman V, Bamberger P A, Wang M, et al. Best not to know: Pay secrecy, employee voluntary turnover, and the conditioning effect of distributive justice[J]. *Academy of Management Journal*, 2021, 64(2): 482–508.
- [31] Jex S M. *Stress and Job Performance: Theory, Research, and Implications for Managerial Practice*[M]. California: Sage Publications Ltd, 1998.
- [32] Hobfoll S E. Conservation of resources: A new attempt at conceptualizing stress[J]. *American Psychologist*, 1989, 44(3): 513–524.
- [33] Guerranti F, Dimitri G M. A comparison of machine learning approaches for predicting employee attrition[J]. *Applied Sciences*, 2022, 13(1): 267–275.
- [34] Gerstner C R, Day D V. Meta-Analytic review of leader-member exchange theory: Correlates and construct issues[J]. *Journal of Applied Psychology*, 1997, 82(6): 827–844.
- [35] Marsh R M, Mannari H. Organizational commitment and turnover: A prediction study[J]. *Administrative Science Quarterly*, 1977, 22(1): 57–75.
- [36] 许为民, 杨阳. 我国员工离职影响因素的研究综述[J]. *东华大学学报(社会科学版)*, 2011, 11(4): 261–268.
- Xu Weimin, Yang Yang. A review of influencing factors affecting employee turnover in China[J]. *Journal of Donghua University (Social Science)*, 2011, 11(4): 261–268. (in Chinese)
- [37] Schaufeli W B, Bakker A B. Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study[J]. *Journal of Organizational Behavior*, 2004, 25(3): 293–315.
- [38] 谢小云, 左玉涵, 胡琼晶. 数字化时代的人力资源管理: 基于人与技术交互的视角[J]. *管理世界*, 2021, (1): 200–216.
- Xie Xiaoyun, Zuo Yuhan, Hu Qiongjing. Human resources management in the digital era: A human-technology interaction lens[J]. *Journal of Management World*, 2021, (1): 200–216. (in Chinese)
- [39] 王聪, 易希薇, 张志学. 大数据时代的管理研究新范式: 以 CEO 解聘问题为例[J]. *管理科学学报*, 2023, 26(5): 200–213.
- Wang Cong, Yi Xiwei, Zhang Zhixue. A new paradigm of management research in the big data era: Taking CEO dismissal as an example[J]. *Journal of Management Sciences in China*, 2023, 26(5): 200–213. (in Chinese)
- [40] Al-Darraj S, Honi D G, Fallucchi F, et al. Employee attrition prediction using deep neural networks[J]. *Computers*, 2021, 10(11): 141–152.
- [41] 任之光, 李金, 赵海川, 等. 大数据驱动的管理决策: 研究范式与发展领域[J]. *管理科学学报*, 2023, 26(8): 152–158.
- Ren Zhiguang, Li Jin, Zhao Haichuan, et al. Big data-driven management and decision sciences: The research paradigm and directions[J]. *Journal of Management Sciences in China*, 2023, 26(8): 152–158. (in Chinese)

Algorithm-driven employee turnover prediction and intervention: A study based on OW property company

*ZHU Bao-quan*¹, *ZHANG Zhi-xue*^{2*}, *WAN Wen*³, *CHEN Jia-mu*¹

1. Onewo Space-Tech Service Co., Ltd, Shenzhen 518049, China;

2. Guanghua School of Management, Peking University, Beijing 100871, China;

3. Faculty of Business and Economics, University of Hong Kong, Hong Kong 999077, China

Abstract: Under the current difficult employment situation, many companies are facing the problem of employees' turnover. Identifying employees turnover intentions, understanding the underlying causes, and implementing preemptive interventions are pivotal strategies to mitigate such intentions. Although prior studies have delved into the factors precipitating turnover intentions, their findings lack robust predictive power. This study aims to examine the determinants of employee turnover intention in real-world corporate settings, with the goal of enhancing firms' capability to predict such intentions and address this issue. This study first identifies imperative predictive variables through feature engineering and employs machine learning algorithms. It then leverages the random forest algorithm to determine the hierarchy of turnover determinants and develop a forecasting model. Meanwhile, a quasi-experimental approach is utilized to validate the predictive model in a corporate setting. Finally, the predictive model is applied to test its validity again across the entire company. Moreover, this research tests the correlation between turnover features and turnover intention and finds that these relationships are consistent with previous studies. This research not only comprehensively explores the influencing factors of employee turnover intention using machine learning but also provides useful advice for companies to reduce employee turnover by intervening early and expressing concern in advance.

Key words: machine learning; random forest; turnover intention; field experiment; intervention