

# 融合心理属性的家政员工隐性缺勤预测研究

曹艳霞, 刘峰涛\*

东华大学旭日工商管理学院, 上海

收稿日期: 2023年12月2日; 录用日期: 2024年1月2日; 发布日期: 2024年1月10日

## 摘要

本文提出融合心理属性的家政服务员隐性缺勤预测模型, 以Y企业作为研究平台, 设计心理量表进行问卷调查, 基于心理弹性理论、心理契约理论和职业倦怠理论构建家政服务员心理属性, 进行相关性分析与心理属性筛选, 确定了用于预测家政服务员隐性缺勤的心理属性。结合企业内数据, 进一步构建XGBoost家政服务员隐性缺勤预测模型, 将基于基础属性和行为属性的模型与融合心理属性的模型预测结果进行对比。结果表明, 融合心理属性的预测模型在准确率、精确率、召回率和F1值等指标上都优于基于基础属性和行为属性的预测模型, 其中预测准确率达到了93.78%, 证明了融合心理属性的预测模型的优越性。

## 关键词

XGBoost, 隐性缺勤, 心理属性, 机器学习

## A Study on the Prediction of Implicit Absenteeism of Domestic Employees by Integrating Psychological Attributes

Yanxia Cao, Fengtao Liu\*

Glorious Sun School of Business and Management, Donghua University, Shanghai

Received: Dec. 2<sup>nd</sup>, 2023; accepted: Jan. 2<sup>nd</sup>, 2024; published: Jan. 10<sup>th</sup>, 2024

## Abstract

This article proposes a prediction model for implicit absenteeism of domestic helpers that inte-

\*通讯作者。

grates psychological attributes. Using Y Enterprise as the research platform, a psychological scale is designed for questionnaire survey. Based on the theory of psychological resilience, psychological contract theory, and occupational burnout theory, the psychological attributes of domestic helpers are constructed, and correlation analysis and psychological attribute screening are conducted to determine the psychological attributes used to predict implicit absenteeism of domestic helpers. Based on data within the enterprise, further construct an XGBoost model for predicting implicit absenteeism of domestic service staff, and compare the prediction results of the model based on basic and behavioral attributes with the model integrating psychological attributes. The results show that the prediction model integrating psychological attributes is superior to the prediction model based on basic and behavioral attributes in terms of accuracy, accuracy, recall, and F1 value, with a prediction accuracy of 93.78%, proving the superiority of the prediction model integrating psychological attributes.

## Keywords

XGBoost, Implicit Absenteeism, Psychological Attributes, Machine Learning

Copyright © 2024 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

近年来,我国家政服务业蓬勃发展,在互联网时代背景下,也迎来了新的转型和升级,服务范围逐步扩大,服务内容更加细化和专业化,人们对家政服务的需求不断增加,同时对家政服务员的服务质量和服务水平也提出了更高的要求。家政服务人员面临新的工作内容与挑战,但由于其职业特殊性,家政员工的社会重视程度不高,缺乏工作成就感,流动性高,平台对员工监管难度大,极易出现员工虽然出勤但是工作效率低下的隐性缺勤现象,进而导致服务质量参差不齐、用户信任度降低等,家政行业的发展很大程度上取决于服务人员的服务质量与职业素养,隐性缺勤问题是家政行业谋求长久与良性发展的过程中亟待解决的问题。

在目前“互联网+家政”的服务模式下,家政服务业呈现平台化、线上运营和去门店化的大趋势,从管理角度来看,家政员工多在家庭场景中提供多种静态服务和动态服务,包括室内清洁、老人陪护、整理收纳等服务内容,其工作地点不固定,大部分时间与管理者处在不同空间内,管理者无法及时了解每位员工的工作状态和服务质量,也就很难及时了解员工隐性缺勤状态从而加以干预。隐性缺勤是人力资源领域研究的焦点之一,目前其研究对象多集中于教师、医务人员和大型企业职工,缺乏对家政行业员工隐性缺勤行为的进一步研究。

针对以上问题,本文设计心理量表,结合Y企业真实数据,基于心理理论选取心理属性,通过相关性分析确定用于家政员工隐性缺勤预测的心理属性,选择XGBoost算法构建家政员工隐性缺勤预测模型,将使用基础属性和行为属性的模型预测结果与进一步融合了心理属性的模型预测结果进行对比,验证融合心理属性模型对家政员工隐性缺勤的预测效果的优越性,提供家政服务行业进行员工隐性缺勤行为的新预测方法,帮助企业及时判断员工状态,采取相应措施以减少企业损失,保证家政行业工作的服务质量和效果。

## 2. 文献综述

### 2.1. 隐性缺勤研究

#### 2.1.1. 隐性缺勤概念

隐性缺勤的概念最早是由英国的心理学家 Cary Cooper 教授在 20 世纪 90 年代提出, 他将隐性缺勤描述为“人由于生病或长时间工作致使工作效率不高时, 应该在家休息, 但却仍然工作”的情况[1], 目前国内外的学者对隐性缺勤的概念还未达成统一, 主要可以概括为两类, 第一类定义为员工身体不适但是仍然参加工作的行为, 第二类定义为虽然身体感到不舒服但仍然选择参加工作, 从而导致工作投入减少和工作效率低下的行为[2]。本文定义的隐性缺勤是一种工作投入减少、完成工作量不高的工作状态, 主要出现在员工生病、压力大或在岗但有事不专心的情况。

#### 2.1.2. 隐性缺勤测量

隐性缺勤的测量主要采取自我报告型的调查方法, 国外目前已有的较为成熟的测量工具包括《工作受限问卷》《斯坦福隐性缺勤简式量表》和《工作效率和活动受损问卷》等, 应用较为广泛的量表有《斯坦福隐性缺勤简式量表(SPS)》和《工作受限问卷(WLQ)》等, 其信度和效度都较高。国内学者对一些较为成熟的测量工具进行了汉化研究, 进行了引进和翻译, 验证中文版问卷具有良好的信度和效度, 同时, 在测量工具研发方面, 国内学者依据实际情况针对我国高新技术企业知识工作者研制出了专门的隐性缺勤测量工具[3]。

### 2.2. 预测模型研究

根据已有文献, 国内外学者在员工行为预测、财务金融预测、灾害预测等多个方面都进行了预测研究, 在员工行为预测方面, 主要对员工不安全行为、员工离职、旷工等行为进行预测研究, 从不同方面为员工管理提供决策依据。

在预测方法方面, 最初学者主要采取统计学的方法进行研究, 从单一变量模型到多元变量分析模型, 提高了预测的准确率, 但是仍旧存在着预测结论的时效性不强等问题, 基于此, 学者将分析概率模型引入到预测模型研究当中, 进一步提高了预测模型的使用范围和预测准确率。近年来计算机技术发展迅速, 人工智能、机器学习等人工智能技术逐渐被应用到预测模型研究当中, 多个研究结果表明使用人工智能的预测模型准确率有所提高, 此外, 将统计学方法与人工智能技术相结合也能提高模型的准确率, 如任婷婷将代价敏感支持向量机作为改进 AdaBoost 算法的基分类器, 建立动态了不平衡预警模型[4]。

## 3. 模型方法

### 3.1. XGBoost 算法原理

XGBoost 是一种经过改进的梯度提升决策树(GBDT)算法[5], 构造决策树(子分类器)后一个子分类器对前一个子分类器的残差进行修正, 各个子分类器结果求和构成总分类器, 比起其他模型, 能够更好地控制过拟合[6], 其本质是不同的单个决策树的组合。假设一个包含  $n$  个样本  $m$  个特征的数据集  $|D| = \{(x_i, y_i)\}$ ,  $|D| = x_i \in R^m, y_i \in R$ , 其中  $x_i$  表示第  $i$  个样本的特征向量,  $y_i$  代表第  $i$  个样本的标签, 则由  $K$  棵树集成的模型输出为:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

XGBoost 模型的目标函数为:

$$Obj(\theta) = \sum_i^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

其中式子右边第一部分为度量预测值与真实值之间的损失函数, 第二部分表示对模型复杂度的惩罚项(正则项), 用来控制模型的复杂程度, 防止过拟合。确定一棵 CART 树需要两部分, 首先是树的结构, 这个结构将输入样本映射到一个确定的叶子节点上, 记为  $f_k(x)$ , 第二部分是各个叶子节点值, 输出的叶子节点序号记为  $q(x)$ , 对应叶子节点序号的值记为  $w_{q(x)}$ , 由定义可得:

$$f_k(x) = w_{q(x)} \quad (3)$$

XGBoost 对应的模型包含多棵树, 每棵树的复杂度定义为:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (4)$$

其中  $T$  为叶子节点的个数,  $\|w\|$  为叶子节点向量的模,  $\lambda$  为正则化惩罚项系数,  $\gamma$  为最小训练损失函数下降值。

根据式子(2), 当第  $t$  棵树生成的时候, 目标函数可表示为:

$$Obj^{(t)} = \sum_{i=1}^n L\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \text{Constant} \quad (5)$$

使用二阶泰勒展开对目标函数进行近似定义, 泰勒展开式为:

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)\Delta x^2 \quad (6)$$

假设令损失函数为泰勒公式中的  $f$ , 令损失函数中  $\hat{y}_i^{(t-1)}$  为泰勒公式中的  $x$ , 令损失函数中  $f_t(x_i)$  项为泰勒公式中的  $\Delta x$ , 则目标函数可以表示为:

$$Obj^{(t)} \approx \sum_{i=1}^n \left[ L\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i \cdot f_t(x_i) + \frac{1}{2} h_i \cdot f_t(x_i)^2 \right] + \Omega(f_t) + \text{Constant} \quad (7)$$

其中,  $L\left(y_i, \hat{y}_i^{(t-1)}\right)$  表示前  $t-1$  棵树组成的学习模型的预测误差, 为常数项, 对优化目标不产生影响, 因此可以将常数项合并去除, 得到目标函数为:

$$Obj^{(t)} = \sum_{i=1}^n \left[ g_i \cdot f_t(x_i) + \frac{1}{2} h_i \cdot f_t(x_i)^2 \right] + \Omega(f_t) + \text{Constant} \quad (8)$$

其中

$$g_i = \text{tial}_{\hat{y}_i^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right) \quad (9)$$

$$h_i = \text{tial}_{\hat{y}_i^{(t-1)}}^2 l\left(y_i, \hat{y}_i^{(t-1)}\right) \quad (10)$$

$g_i$  表示预测误差对当前模型的一阶导函数值,  $h_i$  表示预测误差对当前模型的二阶导函数值, 结合式(3)和式(4), 将  $f_k(x)$  与  $\Omega(f)$  的表达式带入到目标函数中, 得到表达式为:

$$Obj^{(t)} = \sum_{i=1}^n \left[ g_i \cdot w_{q(x_i)} + \frac{1}{2} h_i \cdot w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (11)$$

式(11)第一部分为对所有训练样本进行累加, 因为所有样本都是映射为树的叶子节点, 从叶子节点出发, 对所有的叶子节点进行累加, 可得:

$$Obj^{(t)} = \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (12)$$

令

$$\sum_{i \in I_j} g_i = G_j, \sum_{i \in I_j} h_i = H_j \quad (13)$$

$G_j$  表示映射为叶子节点  $j$  的所有输入样本的一阶导之和,  $H_j$  表示映射为叶子节点  $j$  的所有输入样本的二阶导之和, 可得:

$$Obj^{(t)} = \sum_{j=1}^T \left[ G_j w_j + \frac{1}{2} (H_j + \lambda) w_j^2 \right] + \gamma T \quad (14)$$

其中, 对于第  $t$  棵 CART 树的某一个确定结构而言, 其叶子节点是相互独立的,  $G_j$  和  $H_j$  都是确定量, 因此, 式(15)可以看成是关于叶子节点的一元二次函数。此时目标函数最优值  $w$  为:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \quad (15)$$

将  $w_j^*$  带入到目标函数中, 可得:

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \left( \frac{G_j^2}{H_j + \lambda} \right) + \gamma T \quad (16)$$

上式也称为打分函数, 可以用来衡量树结构的质量, 其值越小, 代表结构越好, 用打分函数选择最佳切分点, 从而构建 CART 树, 确定样本特征的所有切分点, 对每一个确定的切分点进行切分, 切分好坏的标准如下:

$$Gain = \frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (17)$$

$Gain$  表示单节点  $Obj^*$  与切分后的两个节点的树  $Obj^*$  之差, 遍历所有特征的切分点, 找到最大  $Gain$  的切分点即是最佳分裂点, 根据这种方法继续切分节点, 得到 CART 树。若  $\gamma$  值设置的过大, 则  $Gain$  为负, 表示不切分该节点, 因为切分后的树结构会变差,  $\gamma$  值越大, 表示对切分后  $Obj$  下降幅度要求越严。

### 3.2. XGBoost 算法优势

XGBoost 算法主要体现在以下五个方面。第一是精度更高, 传统的梯度提升决策树 GBDT 算法在进行优化的时候只使用一阶导数信息, XGBoost 算法则对损失函数进行了二阶泰勒展开, 从而增加了精度, 并且能够自定义损失函数。第二是灵活性更强, 传统的 GBDT 算法的基学习器为 CART, XGBoost 还支持线性分类器。第三是防止过拟合, XGBoost 通过在目标函数当中加入正则项来控制模型的复杂程度, 从而使模型的方差降低, 学习出的模型变得更加简单, 有助于防止出现过拟合的情况。第四是支持列抽样, 能够在防止过拟合的同时减少计算。第五是增加了对缺失值的处理, 对于样本中特征的缺失值, XGBoost 采用的稀疏感知算法可以自动学习出它的分裂方向。基于以上, XGBoost 能够在取得高精度的同时保持较快的速度, 在使用当中能够取得较好的效果。

## 4. 影响家政员工隐性缺勤预测的心理属性构建

### 4.1. 基于心理理论选取心理属性

心理活动不但对个体的身心健康有影响, 同时在其行为和社会适应等方面都发挥着作用[7], 在人力资源管理领域, 心理属性也已经得到了广泛的应用[8]。基于已有研究可知, 心理弹性、职业倦怠和心理契约是影响员工隐性缺勤的重要因素[9], 每个心理维度下的属性能够更详细地说明产生影响的具

体原因, 因此, 结合心理弹性、职业倦怠和心理契约理论完成影响家政员工隐性缺勤预测的心理属性构建。

基于心理弹性、职业倦怠和心理契约理论的概念、内容与维度、理论模型[10]等方面选取心理属性, 在此基础上, 结合对 Y 企业管理者的访谈结果分析, 保证选取心理属性的合理性, 进行心理属性的初步确定, 最终确定的心理属性如表 1 所示。

**Table 1.** Summary of psychological attributes that affect the implicit absence status of domestic service staff

**表 1.** 影响家政服务员隐性缺勤状态的心理属性总结

维度	心理属性
心理弹性	工作控制、解决问题、工作成就感、抗压力、个人抗逆性、寻求同伴支持、平衡工作与家庭
职业倦怠	人际关系、工作效能、工作意义、工作情绪、工作量、疲劳程度
心理契约	培训制度规范性、晋升制度规范性、薪酬制度规范性、奖励与惩罚制度规范性、公司支持、上级支持

## 4.2. 心理属性实验分析

### 4.2.1. 调研对象

Y 企业在职家政服务人员。

### 4.2.2. 实验方法

为了测量心理属性, 同时考虑到 Y 企业家政员工的受教育水平, 本文基于目前较为成熟的心理弹性量表[11]、心理契约量表[12]、职业倦怠量表[13]、隐性缺勤量表[14]进行修订, 将量表的维度与问题与所构建的心理属性相对应。心理弹性量表和职业倦怠量表采用采取 Likert-5 分量表法进行计分, 1 = 从来没有这样、2 = 很少这样、3 = 有时候这样、4 = 经常这样、5 = 几乎每次都这样, 隐性缺勤量表采取 Likert-5 分量表法进行计分, 1 = 完全不同意、2 = 比较不同意、3 = 不确定、4 = 比较同意、5 = 完全同意, 其中后两个条目采取反向计分, 以得分的中位数为截断点, 分为高隐性缺勤和低隐性缺勤状态[15], 心理契约量表采取 Likert-5 分量表法进行计分, 1 = 不符合、2 = 有点不符合、3 = 一般、4 = 有点符合、5 = 符合。

### 4.2.3. 实验信度检验

使用克隆巴赫系数进行量表的信度检验, 一般来说, 当克隆巴赫系数大于 0.6 时, 说明量表的信度较好, 心理弹性量表、心理契约量表、隐性缺勤量表、职业倦怠量表的克隆巴赫系数均大于 0.6, 同时, 各项的删除项后的克隆巴赫系数也没有明显的提升, 说明无需对题目进行修改, 心理弹性量表、心理契约量表、隐性缺勤量表、职业倦怠量表的信度检验结果如表 2 所示。

**Table 2.** Reliability analysis of questionnaire questions

**表 2.** 问卷题目可靠性分析

维度	克隆巴赫系数	基于标准化的克隆巴赫系数	项数
心理弹性	0.811	0.820	9
职业倦怠	0.618	0.625	8
心理契约	0.746	0.784	12
隐性缺勤	0.776	0.784	6

### 4.2.4. 基本信息分析

本次共调查 801 名 Y 企业家政服务员工, 对收集信息整理后, 调查对象的基本信息如表 3 所示。

**Table 3.** Basic information of survey subjects**表 3.** 调查对象基本情况

人口统计特征变量	变量取值	人数	百分比
性别	男	131	16.4%
	女	670	83.6%
星级	1	441	55%
	2	292	36.5%
	3	68	8.5%

### 4.3. 心理属性相关性分析与属性筛选

为了降低所使用心理属性的维度, 本文使用皮尔逊相关系数对心理属性进行相关性分析, 以此进行心理属性的筛选, 皮尔逊相关系数的计算公式为:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \quad (18)$$

其中,  $\text{cov}(X,Y)$  是  $X, Y$  的协方差,  $\sigma_X$  是  $X$  的标准差,  $\sigma_Y$  是  $Y$  的标准差。皮尔逊相关系数的值在  $[-1, 1]$  之间[16], 所得值的绝对值越大, 证明两属性的相关程度越高, 设定皮尔逊相关系数的阈值为 0.5, 即当属性  $X$  与  $Y$  的皮尔逊相关系数的绝对值大于 0.5 时, 属性  $X$  与  $Y$  的相关程度较高, 从而保留属性  $X$ , 删除属性  $Y$ 。

#### 4.3.1. 心理弹性维度下心理属性相关性分析与筛选

在心理弹性维度下选取的心理属性及其代码如表 4 所示。

**Table 4.** Psychological attributes and codes in the dimension of psychological resilience**表 4.** 心理弹性维度下心理属性及代码

维度	心理属性及代码
个人能力	R1: 工作成就感
	R2: 解决问题
	R3: 工作控制
控制	R4: 平衡工作与家庭
	R5: 寻求同伴支持
相信本能	R6: 抗压力
	R7: 抗压力
积极接受改变	R8: 个人抗逆性
	R9: 个人抗逆性

心理弹性维度下各个心理属性之间的皮尔逊相关系数如表 5 所示。

**Table 5.** Pearson correlation coefficient matrix table between various psychological attributes under the dimension of psychological resilience**表 5.** 心理弹性维度下各个心理属性之间皮尔逊相关系数矩阵表

	R1	R2	R3	R4	R5	R6	R7	R8	R9
R1	1								
R2	0.246**	1							
R3	0.343**	0.445**	1						
R4	0.048	0.136**	-0.011	1					

续表

R5	0.203**	0.306**	0.333**	0.245**	1				
R6	0.370**	0.467**	0.443**	0.075*	0.258**	1			
R7	0.292**	0.498**	0.491**	0.096**	0.428**	0.444**	1		
R8	0.292**	0.524**	0.511**	0.082*	0.300**	0.528**	0.582**	1	
R9	0.378**	0.452**	0.562**	0.067	0.322**	0.519**	0.517**	0.620**	1

心理弹性维度下的心理属性筛选流程如下：依据属性 R1 为初始点进行选择，属性 R1 与其余属性的皮尔逊相关系数都小于 0.5，故不进行属性筛选；根据属性 R2 进行选择，属性 R2 与属性 R8 的皮尔逊相关系数大于 0.5，故删除属性 R8，保留属性 R2；根据属性 R3 进行选择，R3 属性 R9 属性的皮尔逊相关系数大于 0.5，故删除属性 R9，保留属性 R3；同理，分别根据属性 R4、R5、R6、R7 进行选择，属性 R4、R5、R6、R7 均与其余任一属性的皮尔逊相关系数都小于 0.5，故不进行属性筛选，按照此步骤完成心理弹性维度下的心理属性筛选。

经过对心理弹性维度下各个心理属性的相关性分析，本文最终选取了 R1、R2、R3、R4、R5、R6、R7 等 7 个心理属性。

#### 4.3.2. 职业倦怠维度下心理属性相关性分析与筛选

在职业倦怠维度下选取的心理属性及其代码如表 6 所示。

**Table 6.** Psychological attributes and codes under the dimension of occupational burnout

**表 6.** 职业倦怠维度下心理属性及代码

维度	心理属性及代码
去人性化	B1: 人际关系 B2: 人际关系
情感耗竭	B3: 工作情绪 - 情绪反应 B4: 疲劳程度 B5: 工作量
个人成就感	B6: 工作意义 B7: 工作情绪 - 情绪控制 B8: 工作效能

职业倦怠维度下各个心理属性之间的皮尔逊相关系数如表 7 所示。

**Table 7.** Pearson correlation coefficient matrix table between various psychological attributes under the dimension of occupational burnout

**表 7.** 职业倦怠维度下各个心理属性之间皮尔逊相关系数矩阵表

	B1	B2	B3	B4	B5	B6	B7	B8
B1	1							
B2	-0.068	1						
B3	0.474**	0.044	1					
B4	0.355**	0.030	0.416**	1				
B5	0.352**	0.012	0.430**	0.404**	1			
B6	0.048	0.267**	0.100**	0.068	0.055	1		
B7	-0.018	0.365**	-0.042	-0.031	-0.035	0.214**	1	
B8	-0.049	0.493**	0.008	0.001	0.007	0.311**	0.622**	1



由于职业倦怠维度下的心理属性筛选流程与心理弹性维度下的心理属性筛选方法相同, 故不再体现具体分析过程, 经过对职业倦怠维度下各个心理属性的相关性分析, 本文最终选取了 B1、B2、B3、B4、B5、B6、B7 等 7 个心理属性。

#### 4.3.3. 心理契约维度下心理属性相关性分析与筛选

心理契约维度下选取的心理属性及其代码如表 8 所示。

**Table 8.** Psychological attributes and codes under the dimension of psychological contract

**表 8.** 心理契约维度下心理属性及代码

维度	心理属性及代码
员工个人发展	C1: 培训制度规范性 C2: 晋升制度规范性
团队支持	C3: 上级支持 - 情感支持 C4: 公司支持 C5: 上级支持 - 解决冲突 C6: 公司支持 C7: 上级支持 - 工作帮助 C8: 上级支持 - 技能培训
薪酬奖励制度	C9: 奖励与惩罚制度规范性 C10: 奖励与惩罚制度规范性 C11: 薪酬制度规范性 C12: 薪酬制度规范性

心理契约维度下各个心理属性之间的皮尔逊相关系数如表 9 所示。

**Table 9.** Pearson correlation coefficient matrix table between various psychological attributes under the dimension of psychological contract

**表 9.** 心理契约维度下各个心理属性之间皮尔逊相关系数矩阵表

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
C1	1											
C2	0.651**	1										
C3	0.441**	0.402**	1									
C4	0.402**	0.437**	0.362**	1								
C5	0.316**	0.343**	0.382**	0.270**	1							
C6	0.062	0.025	0.010	0.068	-0.066	1						
C7	0.277**	0.326**	0.377**	0.157**	0.513**	-0.008	1					
C8	0.405**	0.359**	0.324**	0.276**	0.248**	0.044	0.219**	1				
C9	0.263**	0.312**	0.141**	0.420**	0.307**	0.026	0.217**	0.181**	1			
C10	0.265**	0.334**	0.246**	0.412**	0.267**	0.028	0.225**	0.205**	0.458**	1		
C11	0.006	-0.017	0.009	-0.055	-0.005	0.287**	0.086*	0.047	-0.047	-0.118**	1	
C12	0.382**	0.358**	0.276**	0.547**	0.305**	0.053	0.219**	0.289**	0.448**	0.448**	-0.101**	1

最终选取 C1、C3、C4、C5、C6、C8、C9、C10、C11 等 9 个心理属性。经过对心理弹性、职业倦怠、心理契约三个维度下各个心理属性的相关性分析, 最终选取了 23 个心理属性用于下一步的模型构建。

## 5. 融合心理属性的家政员工隐性缺勤预测模型构建

### 5.1. 对比试验

#### 5.1.1. 数据预处理

本文使用的数据包括 Y 家政企业数据库内的真实数据和使用心理量表所采集的数据, 根据家政员工姓名和门店将以上两类数据进行匹配连接形成数据集。在构建模型之前, 需要对所使用的数据进行预处理, 避免出现数据缺失、数据类型错误等问题。

在经过缺失值处理、异常值处理、数据类型转换等操作后, 形成包含基础属性、行为属性、心理属性和隐性缺勤状态属性等 49 个属性, 801 个样本在内的实验数据集, 按照 7:3 的比例将数据集划分为训练数据集和测试数据集。

#### 5.1.2. 数据属性说明

经过数据预处理后, 本文使用的实验数据集包含 49 个属性、801 个样本。其中, 结果变量为隐性缺勤状态属性, 其他属性为条件变量, 可细分为基础属性、行为属性和心理属性, 以下将对各数据属性进行详细说明。

将隐性缺勤量表得分的中位数作为截点, 将家政人员隐性缺勤状态分为高隐性缺勤和低隐性缺勤状态, 其中, 高隐性缺勤状态 380 人, 低隐性缺勤状态 421 人。

本文使用的家政服务员数据集中的基础属性包括入职年限、拥有技能数量、学历、性别、星级等 8 个属性, 属性名称定义为 J1-J8, 行为属性包括接单量、工资、请假通过次数、请假拒绝次数等 17 个属性, 属性名称定义为 x1-x17。

#### 5.1.3. 对比实验结果分析

首先构建基于基础属性和行为属性的预测模型, 命名为模型 1, 再构建融合心理属性、基础属性和行为属性的预测模型, 命名为模型 2, 采用网格搜索方法进行参数调优, 调优后的参数如表 10 所示。

**Table 10.** Tuning parameter settings

**表 10.** 调优参数设定

模型名称	参数名称	设定值
模型 1	学习率	0.01
	分类器个数	398
	最大树深度	9
模型 2	学习率	0.01
	分类器个数	448
	最大树深度	5

选取准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 值(F1 Score) 4 个指标来进行模型的评价, 采用经过调优后的参数建立 XGBoost 模型, 对比两种模型的预测效果, 详情见表 11。

**Table 11.** Comparison of model effects

**表 11.** 模型效果对比

模型名称	隐性缺勤状态	Precision	Recall	F1 Score	Accuracy
模型 1	0	0.6415	0.5573	0.5965	0.6183
	1	0.6000	0.6807	0.6378	
模型 2	0	0.9735	0.9016	0.9362	0.9378
	1	0.9063	0.9748	0.9393	

由表可知, 融合了心理属性的模型 2 在四项指标上均优于未融合心理属性的模型 1, 从准确率来看, 模型 2 的准确率比模型 1 提高了 31.95%, 证明了融合了心理属性的 XGBoost 模型在预测家政服务员工的隐性缺勤状态方面的性能更好; 从精确率角度来看, 以高隐性缺勤状态为例, 模型 2 的精确率比模型 1 提高了 30.63%, 表明模型 2 所预测的高隐性缺勤状态中预测正确的比例大幅度提高; 从召回率角度来看, 以高隐性缺勤状态为例, 模型 2 的召回率相较于模型 1 提高了 29.41%, 可知模型 2 能够更全面地正确预测处于高隐性缺勤状态的员工, 同时, 模型 2 的 F1 值高于模型 1, 综合来看, 模型 2 的各评价指标均高于模型 1, 在家政员工的隐性缺勤状态预测方面有着良好的性能。

## 6. 结论

本文首先设计心理测量量表对 Y 家政企业员工进行调查, 基于相关的心理理论完成对心理属性的初步选取, 再通过相关性分析进一步筛选所使用的心理属性, 将 Y 企业数据库数据与量表测量数据进行匹配连接, 经过数据预处理后形成所使用的数据集, 将数据集用于 XGBoost 家政员工隐性缺勤状态预测模型的构建, 采用网格搜索方法进行参数调优, 分别构建两个预测模型, 模型 1 为基于基础属性和行为属性的家政员工隐性缺勤行为预测模型, 模型 2 为融合心理属性、基础属性和行为属性的家政员工隐性缺勤行为预测模型, 将两个模型进行对比实验, 采用准确率、精确率、召回率和 F1 值等指标来评价模型的性能, 结果表明, 模型 2 在各个指标上均优于模型 1, 表明了融合了心理属性的 XGBoost 家政员工隐性缺勤状态预测模型的优越性。

本文结合大数据方法, 融合心理属性构建 XGBoost 家政员工隐性缺勤预测模型, 提高了对家政员工隐性缺勤状态预测的准确性, 便于企业及时获取员工状态, 采取相应措施进行管理, 对企业管理有积极的作用。

## 参考文献

- [1] Cser, J.R. (2010) Impact of Participation in a Worksite Wellness Program on Presenteeism: A Quantitative Study of Pennsylvania Workers. Ph.D. Thesis, Capella University, Minneapolis.
- [2] 刘贝妮, 杨河清. 工作场所隐性缺勤行为研究述评[J]. 经济与管理研究, 2016(4): 66-73.
- [3] 盛利. 高新技术企业知识工作者健康相关工作效率低下测量研究[D]: [博士学位论文]. 北京: 中国协和医科大学, 2008.
- [4] 任婷婷, 鲁统宇, 崔俊. 基于改进 AdaBoost 算法的动态不平衡财务预警模型[J]. 数量经济技术经济研究, 2021, 38(11): 182-197.
- [5] 王重仁, 韩冬梅. 基于超参数优化和集成学习的互联网信贷个人信用评估[J]. 统计与决策, 2019, 35(1): 87-91.
- [6] 王晓晖, 张亮, 李俊清, 等. 基于遗传算法与随机森林的 XGBoost 改进方法研究[J]. 计算机科学, 2020, 47(S2): 454-458+463.
- [7] Myers, D. (2012) Social Psychology. McGraw-Hill Education, New York.
- [8] 谢倩芸, 钱晓焯, 李倩. 团队心理资本对团队与个体任务绩效的影响[J]. 中国软科学, 2022(6): 124-132.
- [9] 李婵. 血液透析室护士职业倦怠、心理授权与隐性缺勤的相关性分析——评《血液透析护理实践精讲》[J]. 世界中医药, 2022, 17(24): 3572.
- [10] Bögemann, S.A., Puhmann, L.M.C., et al. (2023) Psychological Resilience Factors and Their Association with Weekly Stressor Reactivity during the COVID-19 Outbreak in Europe: Prospective Longitudinal Study. *JMIR Mental Health*, 10, e46518.
- [11] 卢瑶, 刘逸飞, 仇玉莹, 等. 突发性公共卫生事件中居民情绪调节对心理健康的影响: 心理弹性的中介作用[J]. 中国健康心理学杂志, 2023, 31(10): 1447-1452.
- [12] 房国新. 组织心理契约管理对员工情绪劳动的影响[J]. 合作经济与科技, 2023(24): 100-102.
- [13] Bassam, S., Mohsen, H., Barakat, Z. and Abou-Abbas, L. (2023) Psychometric Properties of the Arabic Version of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS) among Lebanese Dentists. *BMC Oral Health*, 23, Ar-

title No. 451.

- [14] 金瑜, 宋瑰琦, 丁慧, 等. 心理一致感对妇产科护士包容型领导及隐性缺勤的中介效应[J]. 护理学杂志, 2022, 37(17): 66-68.
- [15] 梁馨之, 孙运波, 尤薇, 等. ICU 护士职业倦怠与隐性缺勤的相关性研究[J]. 中国护理管理, 2017, 17(7): 933-937.
- [16] 张河苇. 基于大数据分析的管道缺陷评估及预测方法研究[D]: [博士学位论文]. 北京: 中国石油大学(北京), 2020.