

基于随机森林的食品安全监测预警研究

刘光惠¹, 陈卓超², 鲜思东², 冯苗苗², 鲜智宇³, 李常郡⁴

1. 重庆市市场监督管理局 档案信息中心, 重庆 400014

2. 重庆邮电大学 理学院, 重庆 400065

3. 南开大学 物理学院, 天津 300071

4. 东南大学 吴健雄学院, 南京 211102

摘要:目的 对商户进行食品安全监测预警是我国食品安全监管中的热点问题, 实现高效的食品安全预警与监管。方法 结合重庆地区食品安全的特点, 采用专家评价的方式得到综合评价指标, 提出 AHP-BP 神经网络综合评价, 建立了具有可靠、客观的食品安全综合评价体系, 在此基础上, 动态挖掘食品安全特征指数, 给出食品安全综合画像, 通过建立基于随机森林的重庆食品安全预警监测的模型, 并对重庆市江北区等食品安全数据进行仿真验证。结果 综合评价指标体系更加合理, 预测的准确率、运行时间等相较 XGBoost 算法均有明显提升。结论 机器学习方法有助于建立更完备、合理的食品安全评价体系, 基于随机森林的食品安全预警模型在精确率、AUC、召回率等方面表现更优, 在商户食品安全监管中不仅精度高, 还有很好的鲁棒性。

关键词: 食品安全; 评价体系; 随机森林; XGBoost 算法; 特征指数

中图分类号: TP181; F203 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2025.0005.004

Research on Food Safety Monitoring and Early Warning Based on Random Forest

LIU Guanghui¹, CHEN Zhuochao², XIAN Sidong², FENG Miaomiao², XIAN Zhiyu³, LI Changjun⁴

1. Information Center, Chongqing Market Supervision Administration, Chongqing 400014, China

2. School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

3. School of Physics, Nankai University, Tianjin 300071, China

4. Chien-Shiung Wu College, Southeast University, Nanjing 211102, China

Abstract: **Objective** Conducting food safety monitoring and early warning for merchants is a critical issue in China's food safety supervision. This study aims to achieve efficient food safety early warning and supervision. **Methods** In combination with the characteristics of food safety in Chongqing, comprehensive evaluation indicators were obtained through expert assessment. An AHP-BP neural network comprehensive evaluation method was proposed, and a reliable and objective food safety evaluation system was established. Based on this, dynamic mining of food safety feature indices was conducted to generate a comprehensive food safety profile. A food safety early warning and monitoring model based on random forest was established, and the model was verified through simulation using food safety data from Jiangbei District, Chongqing. **Results** The comprehensive evaluation indicator system is more rational, with significant improvements in prediction accuracy and runtime compared with the XGBoost algorithm. **Conclusion** Machine learning methods contribute

收稿日期: 2023-12-06 **修回日期:** 2024-02-05 **文章编号:** 1672-058X(2025)05-0027-09

基金项目: 国家自然科学基金(61876201); 重庆市市场监督管理局科技项目(CQSKJ2022023); 重庆市研究生教学改革项目(YJG212022); 重庆市研究生课程思政示范项目(YKCSZ23121); 重庆市市场监管局项目(CQSPJG202401)。

作者简介: 刘光惠(1973—), 女, 重庆江津人, 高级工程师, 博士, 从事计算机软件管理及应用研究。

通信作者: 鲜思东(1971—), 男, 四川南充人, 教授, 从事大数据统计分析与智能决策研究。Email: xiansd@cqupt.edu.cn。

引用格式: 刘光惠, 陈卓超, 鲜思东, 等. 基于随机森林的食品安全监测预警研究[J]. 重庆工商大学学报(自然科学版), 2025, 42(5): 27-35.

LIU Guanghui, CHEN Zhuochao, XIAN Sidong, et al. Research on food safety monitoring and early warning based on random forest[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(5): 27-35.

to the establishment of a more comprehensive and rational food safety evaluation system. The food safety early warning model based on random forest performs better in terms of precision, AUC, and recall rate. It not only achieves high accuracy in food safety supervision for merchants but also demonstrates good robustness.

Keywords: food safety; evaluation system; random forest; XGBoost algorithm; feature index

1 引言

随着我国经济水平的不断提高,外出用餐已成为日常生活中的不可缺少的一部分,食品安全问题备受关注,相关部门和国内外研究人员开始解决食品安全监管指标体系的构建问题。目前,海内外对不同领域综合评价模型的研究已经取得了丰富的成果。综合评价的方法主要分为两大类:一是定性的主观评价;二是定量分析。定性方法主要有专家评议法、层次分析法^[1]、模糊数学评价法^[2]、灰色关联评价方法^[3]、粗糙集理论、风险矩阵等。定量方法主要有因子分析法^[4]、主成分法^[5]、线性模型、贝叶斯模型^[6]、决策树法^[7]等。

食品安全管理是确保食品日常安全的理论和管理方法,旨在保护消费者健康,促进食品行业的公平贸易。近年来,一些学者将机器学习引入食品安全的相关领域,以深度学习^[8]、SVM^[9]、CNN^[10]、贝叶斯分类器^[11]为基础,对食品安全进行了研究。食品安全风险预测预警模型可大致分为两类,离散值预测的分类模型和连续值预测的回归模型^[12]。分类模型一般用于预测食品安全中的风险等级或食品安全检测样本的是否合格的状态,直观地展示食品安全检测样本的整体风险状况,并评估其整体风险状态^[13]。回归模型一般用于预测食品安全检测样本的风险值

或合格率,可以评估食品安全检测样本的风险大小或者检测样本的合格程度,使得风险评估结果更为具体^[14],已有学者采用随机森林方法针对核桃^[15]、牛奶^[16]、果蔬^[17]、海底捞等^[18]视角对食品安全进行检测分析或预测。上述文献中的研究方法尚不能同时兼顾食品安全评价指标体系的构建以及基于食安评价体系的预测与预警,也不能对该算法的时间复杂度与精度进行改进与提升,更不能针对一个城市的综合食品安全建立食安指数,并在此基础上进行全面、有效地评价公众的食品安全。

本文针对重庆市阳光餐饮 APP 中数据的特点,建立合理的综合评价体系并对数据的显著性指标进行分析,构建了基于随机森林的食品安全监测预警模型。并对模型的特征选择进行了进一步探讨。结果表明,该模型的训练速度、运行时间、准确率、精准度及其他指标都较为优秀,可为各地的食品监督部门对类似情况商户安全指数判断提供帮助。

2 食品安全综合评价指标体系构建

2.1 食品安全评价指标的重构

2.1.1 指标体系来源与构建

根据重庆市市场监督管理局原有监管制度整理出指标体系框架如图 1 所示。

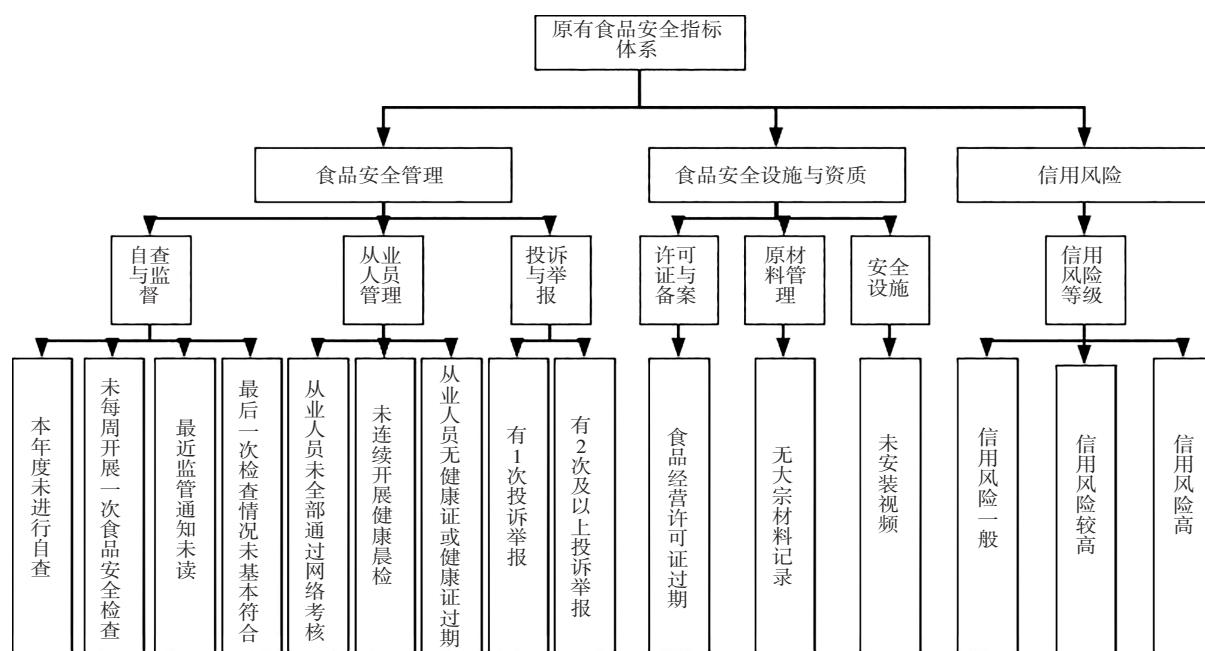


图 1 原有食品安全指标体系

Fig. 1 Original food safety indicator system

由图 1 可得,原有指标体系存在消费者意见以及食品安全事故等方面指标的缺失,指标涵盖不全、不科学。因此,本文首先需要构建一套科学完备的食品安全综合评价指标体系,根据消费者和监管人员进行食品安全评估的客观需要和实践需求等,归纳整理出 3 级指标,然后对其分类,得出 2 级指标和 1 级指标,采用逆向归纳法,重新整理得到的综合评价指标体系如图 2 所示。

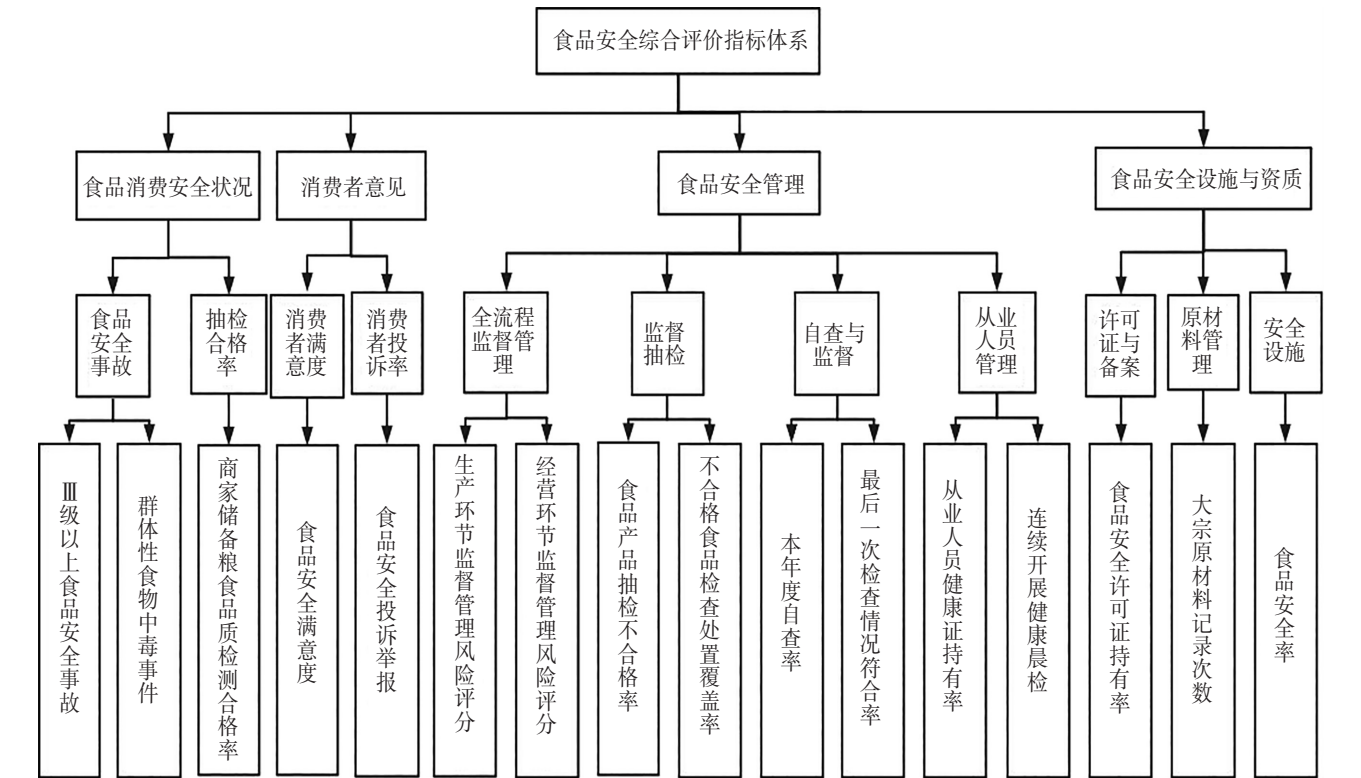


Fig. 2 Comprehensive evaluation index system for food safety

2.1.2 指标体系构建基本理论

(1) 食品安全管理理论。该理论主张食品安全管理对于确保食品质量与安全至关重要,它涉及企业自查、从业人员管理等方面。良好的食品安全管理能降低食品安全事故的风险,提高消费者信心。主要理论来源包括 HACCP(危害分析与关键控制点)体系等。

(2) 食品安全设施与资质理论。该理论强调食品安全设施与资质是企业合法经营和维护食品安全的前提。合规的许可证、备案和安全设施有助于提高企业的信誉度和市场竞争力,主要理论来源包括《食品安全法》、GMP(良好生产规范)等。

(3) 食品消费安全状况理论。该理论指的是消费者在购买和食用食品时所关注的食品质量和安全隐患问题的意识和行为。食品安全事故以及食品抽检合格率严重影响着消费者对商家的评估,主要理论来源包括《食品安全法》、消费者行为学、风险感知理论等。

(4) 社会共享共治理论。该理论侧重消费者对食品安全的评价。消费者的好评与投诉从一定程度上影响着商家的信誉和回头率,主要理论来源包括消费者行为学、社会责任理论。

2.2 AHP-BP 神经网络评价模型优化与应用

层次分析法是由美国运筹学家 Saaty^[19] 于 20 世纪 70 年代初提出,其基本思想是将复杂的决策问题分解为多个层次。该方法能够考虑不同因素之间的相对重要性,通过建立层次结构和构建判断矩阵,可以量化地比较和评估各个因素的重要性,从而更加客观地进行决策。

BP 神经网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家提出的概念^[20]。其核心思想在于运用梯度下降法,通过调整输入节点与隐含层节点的连接强度以及隐含层节点与输出节点的连接强度和阈值,使误差沿着梯度方向逐渐减小。经过反复的学习和训练,使误差降至最小,最终在训练结束时使输出值更加接近目标值。

将原有的指标体系的数据经过 AHP-BP 修改后的综合评价指标及其权重见表 1。表 1 中蓝底为保留原有指标,橙底为新增指标,白底为删减原有指标,本文指标采用归纳法,2 级、3 级指标的理论来源主要基于食品安全领域的相关理论和实践经验,并通过归纳法得出 3 个 1 级指标。

表 1 食品安全综合评价指标体系及其权重

Table 1 Comprehensive evaluation index system and its weights for food safety

一级指标	权 重	2 级指标	权 重	3 级指标	权 重	综合权重
食品安全 管理 A1	0.088 3(后) 0.623 0(前)	自查与监督 B1	0.141 4(后) 0.411(前)	本年度自查率 C1	0.333 3(后)	0.004 2(后)
				最后一次检查情况符合率 C2	0.454 0(前)	0.116 0(前)
				每周食品安全自查率	0.666 7(后)	0.008 3(后)
				最近监管通知查看率	0.323 0(前)	0.083 0(后)
					0(后)	0(后)
					0.072 0(前)	0.018 0(前)
		从业人员管理 B2	0.141 4(后) 0.328 0(前)	从业人员健康证持有率 C3	0(后)	0(后)
				连续开展健康晨检率 C4	0.068 0(前)	0.017 0(前)
				从业人员通过网络考核率	0.666 7(后)	0.008 3(后)
					0.090 0(前)	0.018 0(前)
		全流程监督管理 B3	0.263 0(后) 0(前)	生产环节监督管理风险评分 C5	0.333 3(后)	0.004 2(后)
				经营环节监督管理风险评分 C6	0.767 0(前)	0.157 0(前)
		监督抽查 B4	0.454 7(后) 0(前)	食用产品抽检不合格率 C7	0(后)	0(后)
				不合格食品核查处置覆盖率 C8	0.5(前)	0.011 6(后)
食品安全 设施 与资 质 A2	0.157 5(后) 0.239 0(前)	投诉与举报	0(后) 0.261 0(前)	有 1 次投诉举报	0(前)	0(前)
				有 2 次投诉举报	0.5(前)	0.081 0(前)
		许可证与备案 B5	0.169 8(后) 0.170(前)	食品安全许可证持有率 C9	0(后)	0(后)
					0.5(前)	0.081 0(前)
		原材料管理 B6	0.387 3(后) 0.443 0(前)	大宗原材料记录次数 C10	1(后)	0.026 7(后)
				近一个月大宗原材料记录次数	1(前)	0.041 0(前)
食品 消费 安全 状况 A3	0.484 2(后) 0(前)	食品安全事故 B8	0.833 3(后) 0(前)	Ⅲ级以上食品安全事故发生率 C12	1(后)	0.061 0(后)
				群体性食物中毒事件发生率 C13	0.875(前)	0.093 0(前)
		抽检合格率 B9	0.166 7(后) 0(前)	商家储备粮食品质检测合格率 C14	0(后)	0(后)
					0.125(前)	0.013 0(前)
消费 者意 见 A4	0.271 8(后) 0(前)	消费者满意度 B10	0.666 7(后) 0(前)	食品安全满意度 C15	1(后)	0.069 8(后)
		消费者投诉率 B11	0.333 3(后) 0(前)	食品安全投诉举报次数 C16	1(前)	0.092 0(前)
信用 风险	0(后) 0.137 0(前)	信用风险等级	0(后) 1(前)	信用风险高	0(后)	0(后)
				信用风险较高	0.5(前)	0.069 0(前)
				信用风险一般	0(后)	0(后)
					0.25(前)	0.034 0(前)

3 基于机器学习的食品安全监测预警研究

为更好分析食品安全信息特点及其发展趋势,在对食品安全信息进行综合画像的基础上,本文对该食品安全综合评级得分建立随机森林的监测预警模型,并将其与 XGBoost 算法进行对比,更全面的评价、预测食品安全指数发展规律。

3.1 食品安全信息综合画像

为更好显示食品安全综合情况,本研究选取数据来源重庆市阳光餐饮 APP,具体时间为 2023 年 3 月 1 日—3 月 28 日,共 28 d,数据涵盖了共计 108 576 家商户的信息,包括所有扣分项、加分项以及总得分。通过分析每家商户的日常总体得分,深度挖掘并确定多数商户扣分原因,本文得到数据的综合得分画像,如图 3 和图 4 所示。

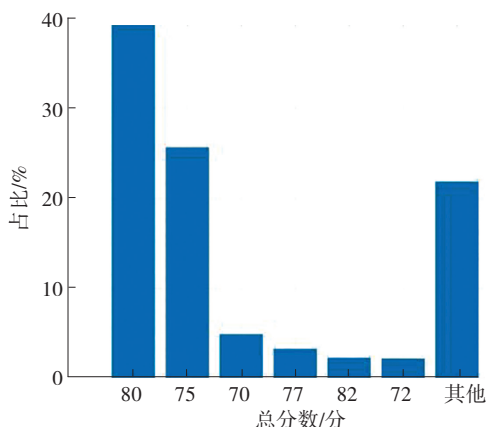


图3 三月份商家总得分情况

Fig. 3 Total scores of merchants in March

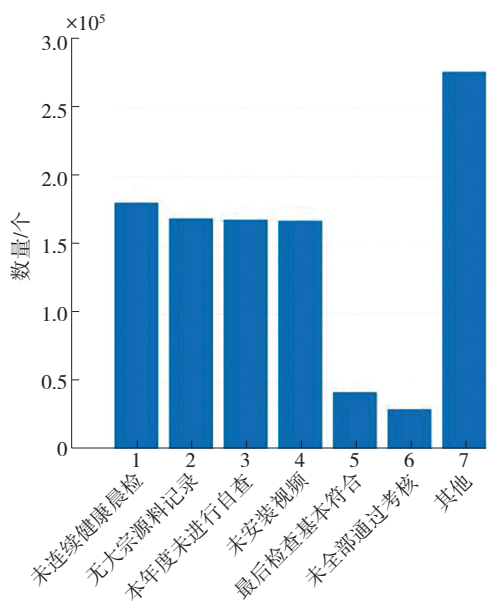


图4 3月28日所有商户的扣分项

Fig. 4 Deduction items for merchants on Mar. 28th

通过图3可得:80分的商家占比最大,75分的商家

次之,两者合计占比65%,且前四项扣分主要原因占比高于73%。进一步分析上述占比最高的两大类商户扣分原因,关键因素为未连续开展健康晨检、无大宗原材料记录、本年度未进行自查和未安装视频。

其次,通过对全部商家的扣分原因进行深度挖掘,结果见图4,排名前4的关键因素仍然为未连续开展健康晨检、无大宗原材料记录、本年度未进行自查和未安装视频。

3.2 随机森林算法的基本原理与步骤

决策树算法是随机森林算法的基分类器,Entropy (熵)代表了决策树系统的表示随机变量不确定性的度量,本文选择使用 ID3, C4.5 和 C5.0 等生成树算法得出熵^[21]。

在分类问题中,假设分类的结果一共有 K 个类别,出现第 k 个类别的概率为 p_k ,则基尼系数的表达式为^[22]

$$G(p) = \sum_{k=1}^K p_k(1-p_k) = 1 - \sum_{k=1}^K p_k^2 \quad (1)$$

对于二分类问题,本文也更能更容易地计算出结果,假设样本中属于第一类输出的概率是 p ,则基尼系数的表达式为^[23]

$$G(p) = 2p(1-p) \quad (2)$$

对于给定的样本 D ,假设有 K 个类别,第 k 个类别的数量为 C_k ,则样本 D 的基尼系数表达式为

$$G(D) = 1 - \sum_{k=1}^K \left(\frac{|C_k|}{|D|} \right)^2 \quad (3)$$

特别地,对于样本 D ,如果根据特征 A 的某个值 a ,把 D 分成 D_1 和 D_2 两部分,则在特征 A 的条件下, D 的基尼系数表达式为^[24]

$$G(D, A) = \frac{D_1}{D} G(D_1) + \frac{D_2}{D} G(D_2) \quad (4)$$

但是决策树算法为弱分类器,易出现过拟合现象,削弱模型的泛化能力。随机森林算法是 Bagging 算法的一种扩展,其基学习器为决策树模型^[25]。为了构建出更具多样性的决策树,在随机森林的训练过程中引入随机性。随机森林的随机性主要分为两重随机:第一重随机是在特征抽取过程中的随机,第二重随机是数据集选择的随机^[26]。

随机森林的预测过程与决策树类似,它通过从根节点开始,依次遍历每个节点,并根据节点的特征值进行判断,从而得到最终的预测结果。在分类问题中,随机森林通常采用投票的方式进行,即采用少数服从多数或者取众数的方式进行最终的预测^[27]。随机森林的具体步骤如下:

(1) 随机抽样训练决策树。根据综合画像得到扣

分量最多的 4 个扣分项作为 4 个最优特征,随机抽取这 4 个特征对应的样本;

(2) 调参、寻找最优特征。向这 4 个特征中不断增添不同数量的新特征,运行结果;

(3) 根据不同参数的 AUC 与性状数的关系、准确率、精确率的大小,选择整体效果最好的特征和参数。

3.3 基于机器学习的食品安全预测分析

3.3.1 数据来源及预处理

本文使用到的数据集为重庆市市场监督管理局 2023 年 3 月 1 日至 28 日在重庆市各个县/区共 104 万商户的食安指数信息,包括商户 ID、住址、总分、加分和扣分明细。总得分是指商家本月按照每日打分和各种加分扣分项目加分的总和。

为防止出现过拟合现象,首先选择了江北区、北碚区、巴南区的商家共计 142 361 户作为初步的数据集。

将商户的扣分项发生次数分为两类,发生记为 1,不发生记为 0。同样,将出现同一类扣分项但仅人数不同的情况,也被认为是同一种。例如,5 位从业人员无健康证或健康证过期与 1 位从业人员无健康证或健康证过期扣分相同。

3.3.2 特征选择

鉴于数据集出现的扣分项较多,易造成过拟合现象。因此,进行有效特征选择至关重要。根据 3.1 节对这些数据的综合画像,选择扣分最多的原因能更轻松地选择到最优的特征。选取 4 个最重要的扣分性状:未连续开展健康晨检、无大宗原材料记录、本年度未进行自查和未安装视频。

3.3.3 划分训练集和测试集

随机选取数据集的 90% 划分为训练集,余下数据

集的 10% 划分为测试集。

3.3.4 评价指标

评价指标选取了 AUC、精确率和准确率 3 个指标。AUC 的值等于 ROC 曲线与 X 轴所围成的面积,ROC 曲线则是一种以 TPR 为纵轴,FPR 为横轴,描述模型获得一定正确率所付出成本的曲线。其中,TPR 为真正例率,即样本的真实标签为正例,且最终结果也为正例的样本所占整体的比例。FPR 为假正例率,即样本的真实标签为负例,且最终结果也为正例的样本所占整体的比例。

3.3.5 特征选择的优化

在构建模型时,首先选择了未连续开展健康晨检、无大宗原材料记录、本年度未进行自查和未安装视频这 4 项作为决策树训练的特征。然而,目前尚无法确定选取这 4 项作为决策树训练性状是否能够达到最优效果。为此,可以向这 4 个特征中不断增添特征。以寻找最优的特征数量和特征。由于随机森林算法的基函数是决策树模型,ID3 决策树受到随机因子和决策树的最大深度的影响较大。为了减弱其影响,参数的选择可以选取两种策略去寻找。第一种策略为分别选择 4 个特征和 18 个特征的最优参数作为两种参数,分别记为参数 1 和参数 2。第二种策略则是根据之前的调参经验,性状数与决策树最大深度为之比为 10:3 时效果最好,为此,将这种策略也可以作为 1 种参数,记为参数 3。在这 3 种参数策略的基础上寻找到最优的特征数量和特征。

通过按照上面的三种策略进行最优特征的寻找,最终寻找到不同特征数量和该数量下的特征,得到如下的关于性状数与准确率、AUC、精确率的关系如图 5 所示。

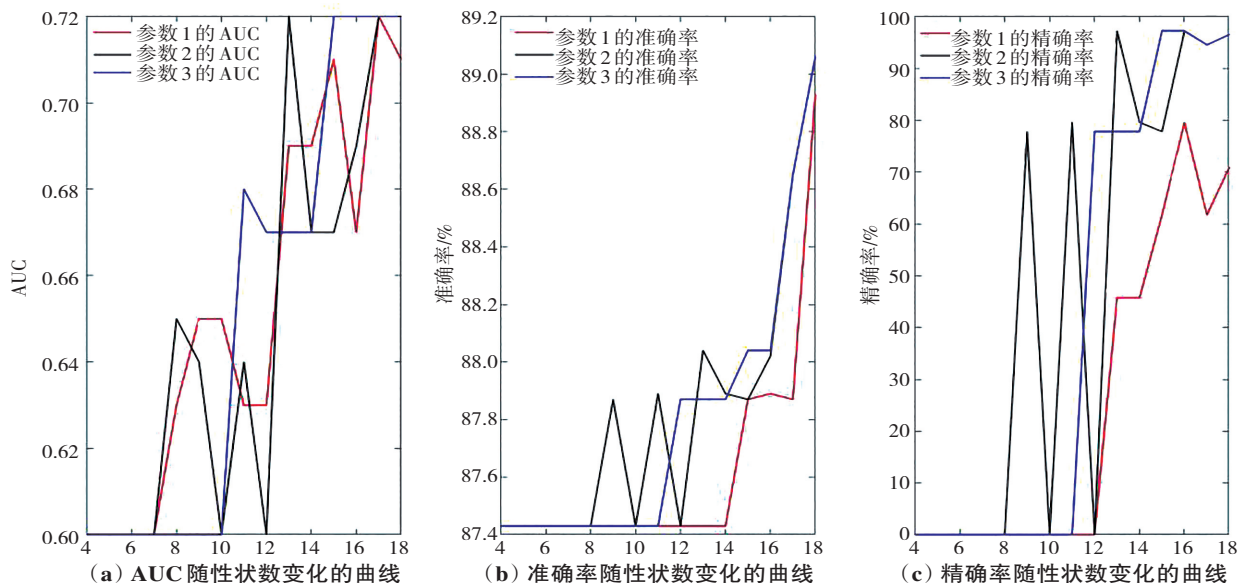


图 5 三种参数下性状数与准确率、AUC、精确率的关系

Fig. 5 Relationship of the accuracy, AUC, and accuracy under three parameters

从图 5 可得,无论是哪种调参策略,准确率、AUC、精确率整体上都会随着特征数的上升而上升。其中特征数为 18 个,调参策略为参数 3 的条件下,准确率、AUC、精确率都是最高的,因此,认为此时的效果最好。

3.3.6 对比研究与结果分析

在训练完随机森林算法的模型的训练后,使用 AUC、准确率、精确率、召回率和模型训练时间 5 个指标,评价其预测性能,并对比 XGBoost 算法在该数据集上的预测性能。XGBoost 算法训练模型依旧采取上述方法,选出 XGBoost 算法对应的最优特征数和特征,以及最优参数,结果如表 2 所示。

表 2 随机森林算法和 XGBoost 算法各项指标的对比
Table 2 Comparison of various indicators between random forest and XGBoost

算 法	AUC	准确率 /%	精确率 /%	召回率 /%	模型运行 时长/s
随机森林	0.71	89.67	98.18	89.58	8.89
XGBoost	0.72	89.06	96.49	87.96	10.03

结果表明:AUC 用于衡量分类器性能,值越接近 1,表示分类器性能越好,随机森林和 XGBoost 都较为优秀;随机森林算法的预测准确率为 89.67%,比 XGBoost 算法的准确率高 0.61%;随机森林的精确率依旧比 XGBoost 算法高;召回率是模型检测到的正确目标的数量占有所有实际目标数量的比例,考虑宁可多预测一些错的也不能漏检,确保每个用户的食安指数的特征都被考虑,从这个维度来看,召回率高的模型考虑得更完善和精确;随机森林模型运行的时间明显比 XGBoost 快很多,说明随机森林的时间复杂度相对较小,表明随机森林算法的时间复杂性比 XGBoost 更优。综上所述,随机森林的预测效果优于 XGBoost 算法。

4 模型的有效性验证

为了评估本文所提出的食品安全预警监测模型的性能和有效性,进行仿真实验。仿真实验使用数据为当月所有商户的信息。为了确保样本具有代表性,采用随机挑选的方式,随机挑选约 3 万个商户和他们的食安信息作为仿真数据集。

在仿真实验中,对随机森林分类器的预警模型进行仿真。通过对测试集的预测,得到以下仿真结果:准确率为 89.67%。这表明所提出的模型在预测商户整改情况方面表现良好。此外,还计算了仿真数据集的精确率为 91.24%,这表明模型能够较准确地识别出需要进行整改的商户。为了更直观地展示模型的性能,本文绘制了 ROC 曲线(见图 6),并根据测试集绘制了混淆矩阵。ROC 曲线的 AUC 值为 0.73,证明所建模型有较好的分类能力。其仿真结果表明其召回率为 89.65%,表明模型在找到那些真正的疑似食品安全问题商户方面具有较好的能力,表明该模型能够较准确的全面估计预测的所有信息,进一步验证了模型的有效性。

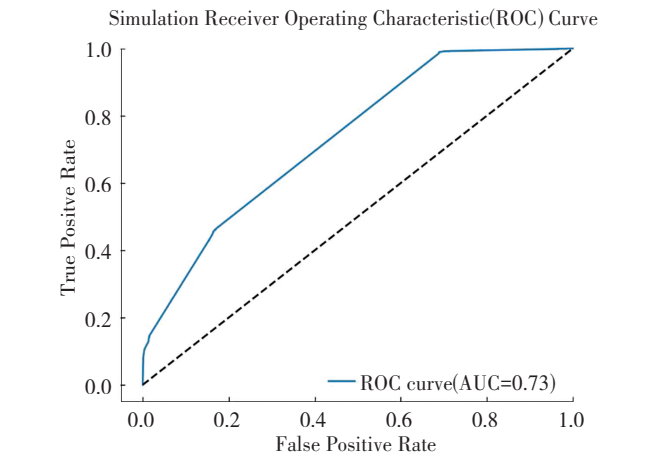


图 6 仿真结果的 ROC 曲线

Fig. 6 ROC curve of simulation results

综上所述,基于仿真的实验结果显示出所提出的模型在预测商户整改情况方面具有较高的准确性和有效性(见图 7)。模型在仿真数据集上的大部分性能与在模型的理论性能相当,甚至在精确率、AUC、召回率 3 个方面还要优于模型在检验集上的结果,证明模型具有较好的鲁棒性和较强的预测性。同时,该模型是基于 AHP-BP 神经网络综合评价模型对多属性数据进行综合画像,建立合理的指标体系,然后进行显著性特征分析,对食品安全指数进行随机森林分类,选出最优的特征和更好的参数用来预测大规模数据,也可以应用于股票、金融、环境影响因素分析、电力预测等领域。

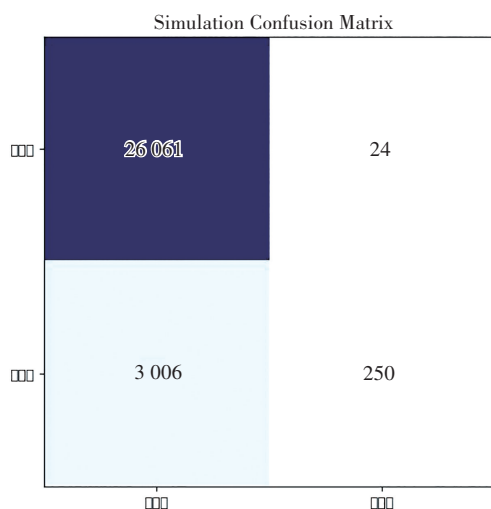


图 7 仿真结果的模糊矩阵

Fig. 7 Fuzzy matrix of simulation results

5 结 论

论文以重庆市市场监督管理局 2023 年 3 月食品安全相关指标数据作为研究对象,构建 AHP-BP 神经网络综合评价模型,并对该数据进行综合画像,在更合理的指标体系之上,又建立一个基于随机森林食品安全监测预警模型,用于判断商户是否存在食品安全问题。首先,使用 AHP-BP 方法改进现有指标体系,其次用综合画像方法选择最重要的 4 个特征,并进行了特征数量和选取特征的优化,以寻找最佳的特征数量和组合。然后,采用不同方式寻找参数组合,优化了随机森林模型。最后,将随机森林模型与 XGBoost 进行了比较。结果表明:随机森林在准确率、精确率、召回率和训练时间等 4 个指标上均优于 XGBoost,表明随机森林模型的效果更佳。随机选用 3 万户商户的数据集进一步仿真验证,结果显示该模型在精确率、AUC、召回率等方面表现更好,在评判食品安全问题商户方面不仅精度高,还有很好的鲁棒性。

未来将会考虑用层次分析法和熵值法相结合以提高初始权值的准确性,并完善成时间复杂度更小、预测效果依旧良好的监测预警体系。

参考文献(References):

- [1] 贾文惠. 基于改进模糊层次分析法的水利工程项目风险评价研究[J]. 陕西水利, 2023(2): 160-163.
JIA Wen-hui. Study on risk assessment of water conservancy project based on improved fuzzy analytic hierarchy process[J]. Shaanxi Water Resources, 2023(2): 160-163.

- [2] XIAN S, LI T, CHENG Y. A novel fuzzy time series forecasting model based on the hybrid wolf pack algorithm and ordered weighted averaging aggregation operator[J]. International Journal of Fuzzy Systems, 2020, 22(6): 1832-1850.
- [3] 王柳璿, 郭焕佳, 姬生国. 基于熵权法和灰色关联度法的铁皮石斛质量评价[J]. 中成药, 2023, 45(2): 483-487.
WANG Liu-ying, GUO Huan-jia, JI Sheng-guo. Quality evaluation of Dendrobium officinale based on entropy weight method and grey correlation degree method[J]. Chinese Traditional Patent Medicine, 2023, 45(2): 483-487.
- [4] XUE WEI, CAI CHAO. Comprehensive evaluation of high-quality development in China based on multi-level factor analysis[J]. Statistics and Decision-Making, 2022, 38(18): 22-25.
- [5] 张柳柳, 黄毓. 基于主成分分析和熵权 TOPSIS 法的医疗质量综合评价研究[J]. 现代医药卫生, 2023, 39(6): 1067-1070.
ZHANG Liu-liu, HUANG Yu. Study on comprehensive evaluation of medical quality based on principal component analysis and entropy weight TOPSIS method[J]. Journal of Modern Medicine & Health, 2023, 39(6): 1067-1070.
- [6] SUN J, WANG F, LI Z, et al. A new hybrid copula-based nonparametric Bayesian model for risk assessments of water inrush[J]. Quality and Reliability Engineering International, 2022, 38(4): 1957-1976.
- [7] 张越, 宋炜炜. 基于 BP 神经网络和决策树的昆明市东川区滑坡空间易发性评价[J]. 国土与自然资源研究, 2023(2): 67-70.
ZHANG Yue, SONG Wei-wei. Spatial susceptibility evaluation of landslide in Dongchuan district of Kunming based on BP neural network and decision tree[J]. Territory & Natural Resources Study, 2023(2): 67-70.
- [8] 李亚军. 基于深度学习的食品安全网络舆情分析[D]. 天津: 天津科技大学, 2018.
LI Ya-jun. Food safety network public opinion analysis based on deep learning[D]. Tianjin: Tianjin University of Science & Technology, 2018.
- [9] 蔡照鹏, 徐林. 易腐食品安全预警云体系构建研究[J]. 河南城建学院学报, 2018, 27(6): 64-70.
CAI Zhao-peng, XU Lin. Early warning method of fresh chicken preservation period based on SVM[J]. Journal of Henan University of Urban Construction, 2018, 27(6): 64-70.

- [10] LIU Y, PU H, SUN D W. Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices[J]. Trends in Food Science & Technology, 2021, 113: 193–204.
- [11] BOUZEMBRAK Y, MARVIN H J P. Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A Bayesian Network approach[J]. Food Control, 2019, 97: 67–76.
- [12] BALYANA K, AHUJA S, LILHORE U K, et al. A hybrid intrusion detection model using EGA-PSO and improved random forest method[J]. Sensors, 2022, 22(16): 5986.
- [13] 王雪冬, 张超彪, 王翠, 等. 基于 Logistic 回归与随机森林的和龙市地质灾害易发性评价[J]. 吉林大学学报(地球科学版), 2022, 52(6): 1957–1970.
- WANG Xue-dong, ZHANG Chao-biao, WANG Cui, et al. Geological disaster susceptibility in Helong City based on logistic regression and random forest[J]. Journal of Jilin University (Earth Science Edition), 2022, 52(6): 1957–1970.
- [14] WU P, WU W, YUAN S. Research on consumers' perception of food risk based on LSTM sentiment classification [J]. Food Science and Technology, 2022, 42: e47221.
- [15] 李世亮. 中国七个主产区核桃综合品质差异比较[D]. 北京: 中国林业科学研究院, 2018.
- LI Shi-liang. The comparison of walnut comprehensive quality from seven main production areas in China[D]. Beijing: Chinese Academy of Forestry, 2018.
- [16] 李佳桐. 基于样本扩充的食品安全风险预测方法的研究[D]. 北京: 北京化工大学, 2022.
- LI Jia-tong. Research and application of food safety risk prediction method based on sample expansion[D]. Beijing: Beijing University of Chemical Technology, 2022.
- [17] 凡建. 基于近红外光谱传感器的便携式食品检测器的设计与实现[D]. 南京: 南京邮电大学, 2019.
- FAN Jian. Design and implementation of portable food detector based on near infrared spectroscopy sensor[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2019.
- [18] 刘迪. 食品安全事件中微博转发影响因素研究: 以“海底捞事件”为例[D]. 天津: 天津财经大学, 2018.
- LIU Di. Study on the influence factors of Weibo retweeting in food safety events[D]. Tianjin: Tianjin University of Finance & Economics, 2018.
- [19] SAATY T L, KEARBS K P. The Analytic Hierarchy Process [M]. New York: John Wiley and Sons, 1980: 22.
- [20] RUMELHART DAVID E, MCCLELLAND, JAMES L. Parallel Distributed Processing: Volume 2 [M]. Cambridge, Mass: The MIT Press, 1986: 217–271.
- [21] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32–38.
- FANG Kuang-nan, WU Jian-bin, ZHU Jian-ping, et al. A review of technologies on random forests[J]. Statistics & Information Forum, 2011, 26(3): 32–38.
- [22] 尹华, 胡玉平. 基于随机森林的不平衡特征选择算法[J]. 中山大学学报(自然科学版), 2014, 53(5): 59–65.
- YIN Hua, HU Yu-ping. An imbalanced feature selection algorithm based on random forest[J]. Acta Scientiarum Naturalium Universitatis Sunyatseni, 2014, 53(5): 59–65.
- [23] CHENT, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785–794.
- [24] WANG J, RAO C, GOH M, et al. Risk assessment of coronary heart disease based on cloud-random forest[J]. Artificial Intelligence Review, 2023, 56(1): 203–232.
- [25] DAI Y, KHANDELWAL M, QIU Y, et al. A hybrid metaheuristic approach using random forest and particle swarm optimization to study and evaluate backbreak in open-pit blasting[J]. Neural Computing and Applications, 2022, 34(8): 6273–6288.
- [26] IQBAL M, ZHANG D, JALAL F E. Durability evaluation of GFRP rebars in harsh alkaline environment using optimized tree-based random forest model[J]. Journal of Ocean Engineering and Science, 2022, 7(6): 596–606.
- [27] 张雪欢, 孙剑伟, 赵黛岩. 基于随机森林的卫星快变遥测数据建模[J]. 计算机测量与控制, 2022, 30(11): 213–218.
- ZHANG Xue-huan, SUN Jian-wei, ZHAO Dai-yan. Modeling of fast changing telemetry data of satellite based on random forest[J]. Computer Measurement & Control, 2022, 30(11): 213–218.

责任编辑:陈芳