

基于机器学习算法的物理专业高校毕业生就业预警模型研究与应用

李昱岑¹ 陈航燕¹ 朱广天² 包 雷³

(¹ 苏州大学物理科学与技术学院物理教育研究所, 苏州 江苏 215006;

² 集美大学师范学院, 福建 厦门 361021;

³ 东南大学生物科学与医学工程学院, 东南大学学习科学研究中心, 江苏 南京 211189)

摘 要 如何有针对性地为就业指导提出建设性意见, 提升高校就业率是各高等院校就业指导工作所面临的严峻挑战。本研究采用机器学习算法, 以我国东部地区 N 校物理类专业本科生就业为例, 以学业及非学业因素为特征, 构建毕业生就业预警模型, 获得了较好的就业预警预测效果。模型呈现以大学英语、线性代数、电动力学及热力学统计物理等科目成绩为重要预测指标的学业预警趋势; 而政治面貌、民族、性别、生源地等非学业因素的特征重要性指数远低于学业因素。本研究旨在为高校理工科类专业就业指导工作提升就业预警干预工作的有效性提供新思路、新方法、新对策。

关键词 人工智能; 机器学习; 大学生就业; 就业指导

RESEARCH AND APPLICATION OF EARLY WARNING MODEL FOR EMPLOYMENT OF COLLEGE GRADUATES MAJORING IN PHYSICS BASED ON MACHINE LEARNING ALGORITHM

LI Yucen CHEN Hangyan ZHU Guangtian BAO Lei

(¹ Institute of Physics Education Research, School of Physical Science and Technology, Soochow University, Suzhou, Jiangsu, 215006;

² Teachers College, Jimei University, Xiamen, Fujian, 361021;

³ School of Biological Science & Medical Engineering, Research Center for Learning Science, Southeast University, Nanjing, Jiangsu, 211189)

Abstract Targeted and constructive advice for employment guidance to enhance the employment rate of colleges and universities is a serious challenge for the employment guidance work of all institutions of higher education. This study adopts machine learning algorithms to construct a graduate employment early warning model according to academic and non-academic factors, taking the employment of undergraduates majoring in physics university N in the eastern China as an example, and obtains a satisfactory prediction results of employment early warning. The model presents the trend of academic early warning with the grades of English, linear algebra, electrodynamics and thermodynamics statistical physics as the important features; while the importance index of the features of non-academic factors, such as political landscape, nation, gender, and origin, is much lower than that of academic factors. This study aims to provide new ideas, methods, and countermeasures for the employment guidance

收稿日期: 2024-04-03

基金项目: 国家自然科学基金(项目编号: 11904246); 苏州大学物理科学与技术学院 2022 年度东吴攀登计划教改类项目(项目编号: NH10840022); 教育部人文社会科学研究青年项目(项目编号: 24YJCZH474)。

通信作者: 陈航燕, chenhangyan@suda.edu.cn。

引文格式: 李昱岑, 陈航燕, 朱广天, 等. 基于机器学习算法的物理专业高校毕业生就业预警模型研究与应用[J]. 物理与工程, 2025, 35(3): 142-152.

Cite this article: LI Y C, CHEN H Y, ZHU G T, et al. Research and application of early warning model for employment of college graduates majoring in physics based on machine learning algorithm[J]. Physics and Engineering, 2025, 35(3): 142-152. (in Chinese)

team of science and engineering majors in universities to improve the effectiveness of employment early warning intervention.

Key words artificial intelligence; machine learning; university students' employment; employment guidance

1 高校就业指导困境剖析

1.1 毕业生就业形势日益严峻

受我国高等教育扩招政策影响,高校毕业生数量逐年上涨。据教育部最新统计,2023 年全国高校毕业生人数达到 1185 万人,同比增加 82 万人,增长率达到 7.62%。疫情对于世界整体经济形势的影响尚未完全恢复,企事业单位用人需求不稳定,灵活用工需求增加。整个就业市场正处于恢复期,高校毕业生正在面临日趋增长的竞争压力。同时,就业市场内部的结构性矛盾难以缓和。后疫情时代,岗位需求和就业渠道不断发生改变,制造业和服务业用人紧缺,而大学生对就业的心理预期没有改变,毕业生人才结构与市场需求不相匹配,种种原因造成了毕业生就业形式日益严峻^[1]。

1.2 传统就业指导模式存在局限性

高校传统就业指导服务模式一般包括授课、宣讲、联系用人单位、召开招聘会、提供求职平台等。对高校毕业生提升就业竞争力虽有助力,但由于该模式仅针对求职现状提供帮助,时效性差、同质化求职讲座等问题难以避免。整体而言,就业指导系统无法从“预警”角度切入干预,导致专业差异化、个性化、针对性、匹配性功能无法有效发挥^[2]。而真正有效的就业指导应当摒弃被动而同质化的“产品包装”,将重点转向动态整合学生资源,深度适应学生发展的“产品设计”^[3]。

1.3 高校就业指导工作边缘化

正是由于“预警”模式的就业指导针对性、差异化、个性化等功能的欠缺,就业指导效果欠佳,队伍建设不受重视、缺乏专业性等因素使得职业生涯规划等就业指导服务往往流于形式^[4]。高等院校参与就业的学生在校经历、学业成绩、社会实践与其就业去向密不可分,但传统就业指导系统缺少与高校教务部门、学生社团、实习系统等主体的深度合作,无法充分利用高校丰富的资源为学生提供优质的就业指导服务^[5]。

2 高等院校物理类专业就业预警模型研究缺乏

我国采用机器学习算法进行大学生毕业去向预测研究在知网最早可查于 2008 年。赵波等人对人才就业市场的数据进行挖掘,构建了学生就业贝叶斯网络模型,找出就业受择业观念、能力素质、择业技巧、就业心理等因素影响的相互依赖关系^[6];徐岩等采用支持向量机算法,基于学生个人情况、学习能力、住址环境、经济状况、家长意见等因素对学生毕业选择进行预测,采用十折交叉验证对模型进行评估并讨论核函数和惩罚变量的参数选择^[7]。何玮采用了 K 均值算法、CJIRE 算法、DBSCAN 算法三种聚类算法对高校毕业生就业指导进行了分析并讨论了无监督聚类算法的设计思路^[8]。近年来,采用机器学习算法进行高校就业预测的相关研究热度呈上升趋势。2017 年,孙怡帆等采用逻辑回归算法基于人力资本理论对毕业生去向进行了预测,并就成因进行分析^[9];刘中国等采用决策树、神经网络、贝叶斯网络算法分别对江苏省某高校金融学学生毕业去向进行了预测,并提出相应的建议^[10]。也有国外学者采用机器学习算法对学生就业情况进行研究,K. Sripath Roy 等人采用随机森林、支持向量机、XGBoost 方法研究了学生成绩、兴趣、人际关系、编程能力等因素对学生在计算机科学领域就业情况进行了预测^[11];N. VidyaShreeram 等人采用决策树、随机森林等算法,以学生年龄、父母状况、健康状况等因素为自变量对学生职业生涯进行了预测^[12]。然而,这些研究往往偏重管理学、计算机科学等方向,鲜有分析高校物理类专业毕业生就业去向及就业预警模型的相关研究。不仅如此,已有的研究中对于学生学业成绩——这一衡量高校教学质量十分重要的因素的分析也较为粗糙,往往仅取学生 GPA 或平均成绩进行研究,而缺少学生各科成绩对于就业重要性的对比。高校物理类本科专业以“难学”“难就业”著称,诸多学生存在盲目的就业或考研从众性和盲目性^[13]。此外,就业与学业关系密切;学业以就业为导向,就业以学业为前

提和基础^[14]。即便有再高的就业意向,无法从学业角度满足基本就业需求,脱离学业依托的就业指导往往是无效的。

本研究基于机器学习算法,对学生就业情况的大数据进行早期预警,以高校大学生接受高等教育的前期(大一、大二)学业成绩为切入点,针对性地从可能存在就业困难的学业特征有所警示,从根源上为大学生提供有效的专业化、个性化就业/考研建议,为有效提高就业指导服务的针对性,扩大校方就业干预的时间窗口,进而在一定程度上提升高校就业指导效果。该算法在处理大样本复杂数据方面相较于传统方法更具优势,通过客观反映当前依托学科专业所反馈出的模型预测结果,有效辅助指导教师的主观判断结果,提升就业指导的有效性;同时也能够充分利用学生校内校外的各方面的数据信息对学生就业情况进行综合分析预测,可以有效改善目前就业指导工作与高校其他部门缺乏合作的困境。

本研究重点分析高等教育物理学专业学生的学业及非学业因素对就业的影响,回答两个问题:(1)高校物理学专业影响学生就业去向的学业因素指向哪些学科成绩?(2)哪些非学业因素可能为就业去向提供指导建议?以期为指导教师有针对性的关注学生成绩提供理论依据。

3 就业预警模型构建方法与流程概述

3.1 人工智能与机器学习算法

随着新一轮科技革命的来临,人工智能技术被不断应用于教育领域。2022年科技部等六部门联合印发的《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》中强调了人工智能高水平应用对于中国高质量发展的关键支撑作用,并明确指出应着力挖掘人工智能在教育等领域中的应用,围绕安全便捷智能社会建设打造重大场景。而机器学习作为人工智能技术的一个重要分支,因为其优秀的数据处理能力和对于动态、复杂、大样本问题的强大适应性,在教育领域中有广阔的发展前景^[15]。机器学习在学界有着诸多定义,其中被广泛采纳的定义是:机器学习是计算机不依赖于明确编程而从数据中学习,进而推广实例执行任务的算法。机器学习的核心目标是通过从数据中发现模式和规

律,进行预测、分类、聚类、优化等任务。

机器学习可以分为监督学习、无监督学习和强化学习三种主要类型:在有监督学习中,训练数据集包含输入特征和对应的标签或输出。通过使用这些已知的输入和输出对模型进行训练,使其能够预测新的未知数据的输出。常见的监督学习算法包括线性回归、决策树、支持向量机和神经网络等;在无监督学习中,训练数据集只包含输入特征,没有对应的标签或输出。无监督学习的目标是从数据中发现隐藏的模式和结构,进行聚类、降维、异常检测等任务。常见的无监督学习算法包括K均值聚类、主成分分析和关联规则挖掘等;在强化学习中,模型通过与环境进行交互,通过试错和奖惩机制来学习最优的行为策略。强化学习常用于智能体在复杂环境中做出决策和控制的问题,如机器人控制和游戏智能等。有监督的机器学习技术在教育学领域中的应最为广泛,应用场景主要分为两类,其一是关于基于图像识别和自然语言识别技术的智能教学设备研究,以为师生提供更为智能高效的学习体验;另一研究方向是基于统计学和数据挖掘技术的教育教学分析及评价,可以有效减少教育工作者的重复劳动,同时剔除由于个体认知、思维模式和情感态度带来的评价误差,提高分析及评价的一致性和科学性^[16]。

在高等教育实践中,传统的就业指导模式主要依赖于教师的专业知识和教学经验。然而对于这一大样本、高复杂度、动态性的问题,传统就业指导方式缺乏针对性、科学性,容易陷入经验主导型教学和主观性教学误区。因此,有必要不断引入有监督机器学习算法等智能技术辅助教师进行更加全面而精准的教学干预^[17]。

本研究横向对比了多种有监督机器学习算法对于大学生就业预警的效果,所采用的算法包括:朴素贝叶斯算法、支持向量机算法、K最近邻算法、决策树算法及随机森林算法。各算法关键理论原理见附录。

3.2 构建就业预警模型流程概述

就业指导预警模型建构流程如图1所示:对东部地区N校物理专业17、18、19级共319名学生进行数据采集、脱敏和描述性统计分析;进行数据预处理(缺失值填补、数据无量纲化、分类型数据编码与哑变量处理、训练集测试集划分);对数据各特征相关性与差异性进行分析,选取最佳特

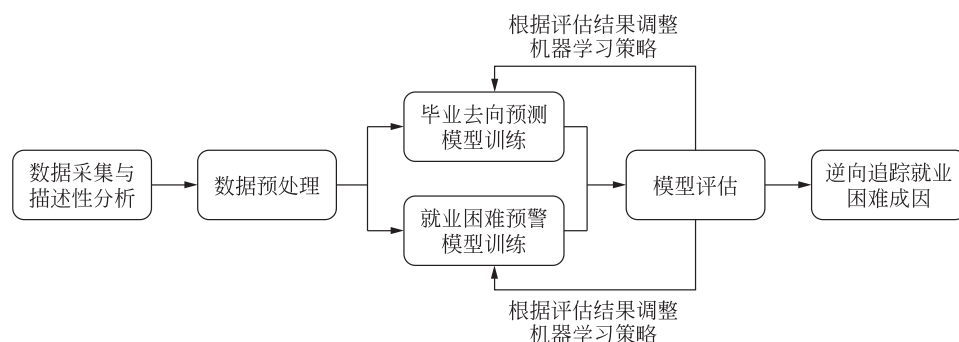


图1 就业预警模型构建流程

征组合；采用决策树、朴素贝叶斯、支持向量机、随机森林、K 最近邻算法构建预测模型，并对毕业去向预测与就业升学困难预警两个目标标签分别进行模型评估；根据评估结果选择最佳模型——随机森林模型调整参数，直至模型效果达到最佳；对整个模型训练过程进行分析，并逆向追踪影响导致学生就业困难的各特征因素。

由于算法自身原理限制以及大学生就业相关数据内禀的样本分布不均衡性等问题，模型对于整体预测的准确率与对于各个目标变量预测的精准度与灵敏度往往难以兼顾。当专注于某一个目标变量的预测效果时，将会导致对于其他目标变量的预测效果变弱，进而导致整体准确率的下降。综合考虑模型特性及就业指导的现实需要，通过调整模型参数来引导机器学习倾向性的方式，将整个模型构建及应用过程划分为三个重要部分：大学生毕业去向预测；大学生就业困难预警；大学生就业困难成因追踪。以期让模型在每个研究部分中都能取得最佳效果。该指导模型将首先重点预测出学生升学或就业的倾向，为就业指导教师对学生进行更具针对性的辅导提供参考依据；其次，模型将尽可能捕捉可能会出现就业升学障碍的学生，并进行预警，以便高校重点关注；最后，模型将返回在模型训练过程的各个特征的重要性，以便在学生呈现出就业困难特征时，及时关注，进行预防。为高校学生提供全方位、系统化的就业指导。

4 数据清洗

4.1 描述性统计

本研究的数据样本来自 N 校的 17、18、19 三个年级共 319 名物理专业大学生。进行数据脱敏

后，分别选择了每名学生的基本信息（包括生源地、性别、政治面貌、民族、城乡户口共 5 项信息）、专业课成绩（包括力学、热学、光学、电磁学、实习成绩、毕业论文成绩等共 14 项科目成绩）、公共课成绩（包括高等数学、大学英语、Python、公共体育、思想道德基础与法律修养、职业生涯规划等共 26 项科目成绩）作为特征变量；毕业去向信息作为标签变量进行建模。所研究学生样本基本信息分布情况如表 1 所示。

表 1 研究样本基本信息分布情况

变量名称及数据内容		人数	所占比例
生源地 (根据生源地 高考试卷 分区分类)	教育较落后地区	33	11%
	教育发展中部地区	47	16%
	教育较发达地区	62	22%
	教育发达地区	145	51%
民族	汉族	264	92%
	其他民族	23	8%
政治面貌	中共党员	9	3%
	其他	268	94%
城乡生源	城市生源	170	59%
	农村生源	116	40%
性别	男	194	68%
	女	93	32%
毕业去向	国外升学	42	13%
	国内升学	137	48%
	待就业	62	22%
	就业	78	27%

学生成绩信息为连续型变量，各科目成绩平均值为 80.68，标准差为 5.75，其正态分布 P-P 图、Q-Q 图如图 2、3 所示，可以认为学生各科目平均成绩整体呈正态分布。

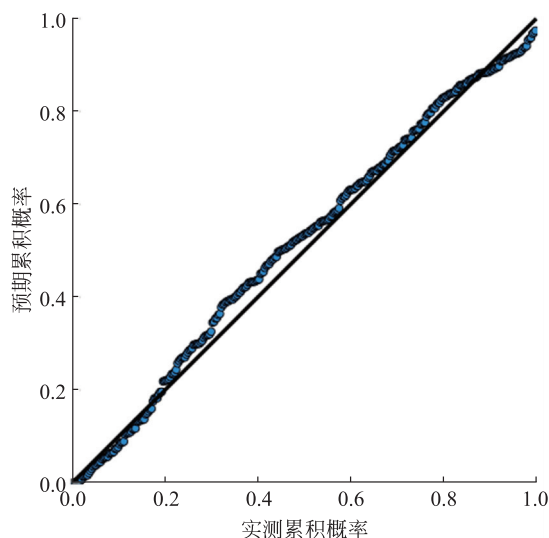


图2 学生样本平均成绩 P-P 图

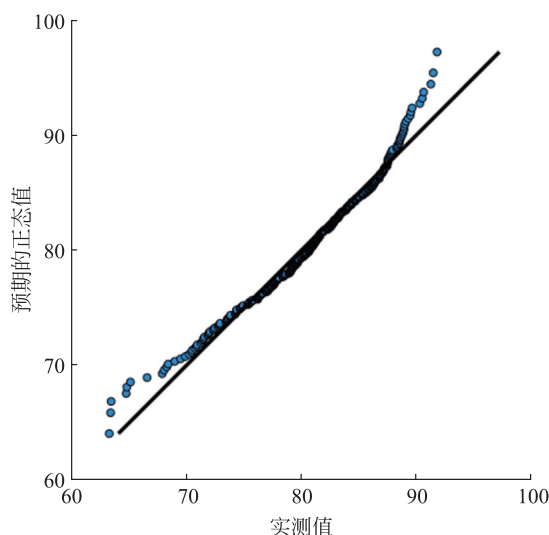


图3 学生样本平均成绩 Q-Q 图

4.2 数据预处理

4.2.1 缺失值填补

由于学生学年不同、选课情况不同等原因,所收集数据中约含 3% 缺失值。缺失数据按变量类型可分为分类型数据(如性别、民族等)和连续型数据(如各科成绩)。本研究使用机器学习算法对缺失值进行预测并填补,该缺失值处理策略相较于传统的均值填补或众数填补,经检验具有更高的准确性。

填补时,对于含有缺失值的各特征,需要按照缺失值从多到少的顺序进行填补。首先从原数据集中将要填补的特征复制提取出来作为目标变量,其余特征作为特征变量构成临时数据集,并根

据变量类型分别采用众数和均值对特征变量进行缺失值填补。其次,将目标变量完整的数据作为训练集,目标变量有缺失的数据作为测试集,使用机器学习算法进行模型训练。接着,再用模型预测出的数据代替缺失值返回原数据集并选取下一项特征进行填补,循环这一过程,直至所有缺失值都被填补完整。

4.2.2 分类型变量编码及哑变量处理

大多数机器学习算法不支持输入文本型数据,需要将文本型数据(如性别男、女等)按照一定规则转化为可供算法识别的数字型数据(如 0、1……),这一过程就是编码过程。

而变量由是否具有数学性质又分为名义变量、有序变量、等距变量、等比变量等。对于无法进行数学运算的名义变量特征(如性别等)需要进行哑变量处理。即将特征“性别”(其中 0 代表男、1 代表女)拆解成两个特征——“性别男”和“性别女”(其中 0 代表是、1 代表否)。这一处理可以让算法将无法进行数学解释的名义变量拆分成若干有序变量进行理解。

4.2.3 连续型变量无量纲化处理

在机器学习算法实践中,往往有着将不同规格的数据转换到同一规格或某个特定分布的需求,这种需求统称为将数据“无量纲化”。在以梯度和矩阵为核心的算法中,如:逻辑回归、支持向量机、神经网络,无量纲化可以加快求解速度;而在距离类模型,如:K 近邻、K-Means 聚类中,无量纲化可以辅助提升模型精度,避免某一个取值范围特别大的特征对距离计算造成影响。在本研究中,选择使用标准化方法对学生百分制成绩数据进行处理(公式 1)。 x^* 为标准化处理后的分数, x 为原始分数, μ 为原始分数平均数, σ 为原始分数标准差。

$$x^* = \frac{x - \mu}{\sigma} \quad (1)$$

5 模型训练

5.1 数据分集

为检验模型效果需要将整个数据集按照 7 : 3 的比例随机划分为训练集(用于训练机器学习模型)与测试集(验证模型预测效果),以供模型进行学习并对学习效果进行评估。

5.2 算法选择

在具体算法的选择上,本研究采用决策树、随机森林、K 最近邻、支持向量机和朴素贝叶斯五种机器学习算法对样本进行预实验模型训练。根据训练结果(图 4),选择模型预测率最高的随机森林算法进行进一步参数调优,以获得最准确的预测结果。

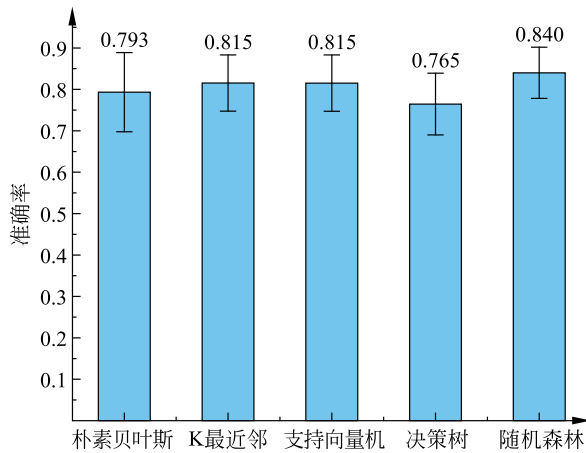


图4 预实验各模型准确率

5.3 模型训练与调参

综合考虑算法固有特征及就业指导实际需要,本研究构建了“物理专业大学生毕业去向预测”三分类与“物理专业大学生就业困难预警”二分类两种分类预测模型并进行参数调优,反向追踪造成学生就业困难的因素。两个分类模型的训练侧重点有所不同,对于毕业去向预测问题,机器学习模型的建构应倾向于最大程度地保证预测的整体准确性;而针对就业困难预警问题,机器学习算法需要尽可能地预测出全部的存在潜在就业问题的样本,即更注重模型的灵敏度。

在训练过程中模型的性能往往受到各种超参数的影响,本研究采用网格搜索法(Grid Search)进行调参。其原理为遍历给定的超参数空间并进行计算,最终找到最优超参数组合。

6 研究结果

6.1 物理专业大学生毕业去向预测

在机器学习相关研究中,常采用混淆矩阵对模型预测结果进行评估。以二分类问题混淆矩阵为例(图 5),矩阵根据样本的真实分类和模型的预测分类被分为 TP、FP、FN、TN 四部分。其中 T

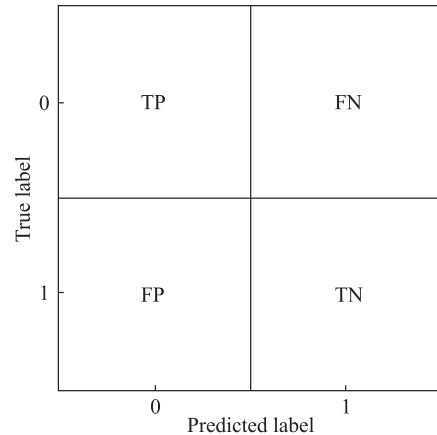


图5 二分类问题混淆矩阵

与 F 代表模型是否判断正确,P 与 N 代表模型判断样本为正或负。为准确具体评估模型的性能,定义了如下评估参数。

① 准确率:模型整体预测正确率。

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

② 精准度:模型反馈正确判断正样本的准确率。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

③ 灵敏度:模型对于正样本的识别、捕捉能力。

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

物理专业大学生毕业去向预测模型测试结果如图 6 所示,模型整体准确率达到了 76.54%,在相似数据集的机器学习案例中位于较高水平,可以有效预测学生毕业去向^[18]。通过进一步分析

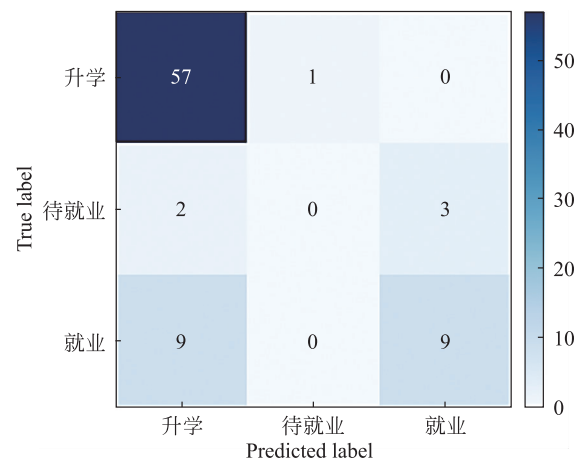


图6 物理专业大学生毕业去向三分类预测结果

可知,模型对于倾向升学的学生有着较强的预测能力,正确率可达到 83.8%,但对于待就业的学生预测的灵敏度较差。导致该结果的原因是原数据中样本不均衡使得升学样本数较多而待就业样本数较少。机器学习模型的预测效果往往随样本量增加而增强,扩大采样规模、获取更多高校毕业生就业数据用于模型训练将有效改善该模型预测效果。

6.2 物理专业大学生就业困难预警

通过前述三分类研究结果分析可知,大学生毕业去向预测模型虽然对学生整体的毕业去向有着较好的预测效果,但对于就业困难学生的鉴别能力较差,而这些学生又是高校就业指导中最需要关注的群体。故本研究将顺利升学、顺利就业的样本合并为一类,将原本的三分类问题转化为二分类问题,重整样本分布不均,以重点识别潜在就业困难个体,凸显就业困难学生特征。

由于算法原理的限制,对于某类样本的预测灵敏度越高,整体的判断准确率就会下降。为实现对于潜在就业困难学生的识别,需要平衡模型整体准确率和对于就业困难学生的灵敏度,通过调整分类阈值,如图 7 所示,可以得到反应模型灵敏度与误判正样本率——假正率(False Positive Rate)二者间关系的 ROC 曲线(模型工作特征曲线)。

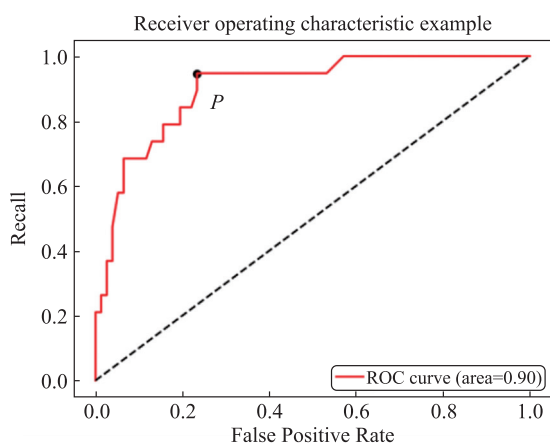


图 7 随机森林模型 ROC 曲线

机器学习算法实践中,常根据实际需求,阅读 ROC 曲线来选择合适的阈值参数建立模型。而 ROC 曲线与 x 轴所围面积称为 AUC 面积,用于量化研究在不同超参数下模型的工作能力,其取值通常在 0.5~1 之间,数值越接近 1 代表模型

在各个阈值参数下都有完备的预测能力,而数值接近 0.5 则代表模型分辨能力弱,近似于随机猜测。

观察 ROC 曲线可以看出,模型灵敏度最初随假正率提高而迅速提高,在 P 点后趋于不变, P 点所对应的参数可以在仅牺牲少量整体准确率的前提下得到最高的灵敏度,故选择 P 点对应的分类阈值进行模型训练。预测结果如图 8 所示,模型准确率为 78.13%,灵敏度为 94.74%,精准度为 50.00%,AUC 值为 0.91。基于此,就业困难预警二分类模型相较于前述三分类模型对于就业困难样本的灵敏度显著提升,仅漏判了一个就业困难样本,满足了在保证整体准确率的前提下对存在就业困难倾向的学生具有较高灵敏度的需求。

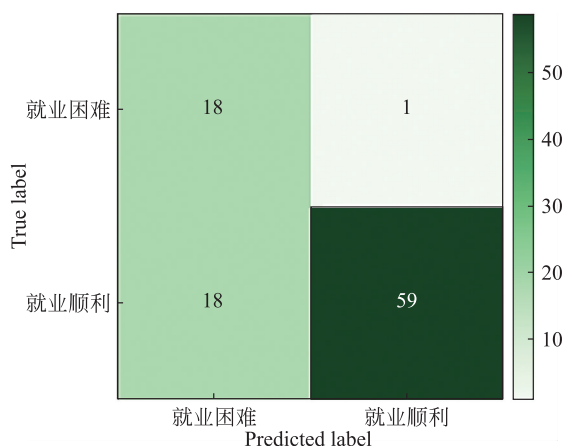


图 8 物理专业大学生就业预警预测二分类结果

6.3 大学生就业困难特征分析

随机森林算法可逆向追踪所有归一化(各特征重要性数值和为 1)的各特征重要性。表 2 呈现了大学生就业预警模型中特征重要性最高的十个成绩类特征与各个非成绩类特征的重要性。其中,英语(大一、大二)、线性代数(大一上)、电磁学(大二上)、电动力学(大二下)、热力学统计物理(大三上)、量子力学(大三上)等课程重要性显著高于其他特征。机器学习算法所判断的重要特征涵盖了数学、物理学等专业核心课程和计算机、英语等通用课程,与基于教育学实践的传统判断高度吻合。这些作为科目成绩可以作为大学生入学早期就业预警信号,相关专业就业指导工作组教师通过对在这些科目表现不佳的学生进行及时干预,提前规避该群体今后可能出现的就业问题。

表 2 各特征重要性排序

特征类型	特征名称	特征重要性
学业成绩	英语均分	0.0722
	线性代数	0.0722
	电动力学	0.0684
	热力学统计物理	0.0563
	量子力学	0.0499
	电磁学	0.0492
	高数(一)上	0.0392
	原子物理学	0.0333
	数学物理方法	0.0327
	计算机信息技术 (计算思维)	0.0301
非成绩因素	城乡户口	0.0052
	生源地	0.0049
	性别	0.0033
	民族	0.0032
	政治面貌	0.0007

城乡户口及生源地这两项常规就业、升学指导工作较为关注的影响因素^[18]是所有非成绩因素中,对大学生能否顺利就业升学影响最大的。然而,相较成绩类因素,非成绩因素特征对于模型预测整体贡献均较低。

对本次研究中 N 校本科生毕业去向数据进行分析发现,除学生个体因素外,社会经济环境和行业发展等宏观因素对于本科生毕业去向同样具有重要影响。根据国家统计局数据,近三年全国城镇调查失业率均高于 5%,仅 2022 年一年全国城镇登记失业人员数量就从 1040 万人激增至 1203 万人,且 2023 年仍有增加趋势^[20]。而全国范围内就业市场的低迷对高校本科生就业情况产生了直接影响,N 校 2021 至 2023 年度本科毕业生选择直接就业的比例持续走低,分别为 40.8%、33.7%、19.0%。

对 N 校 2021 至 2023 年度三届毕业生的从业具体情况进行进一步分析,如图 9 所示,大量本科生选择从事教育行业,且从事教育行业的人数比例逐年升高,直至 2023 年已达 70.8%,而从事制造业、科学研究和技术服务业等其他与物理学科高度相关行业的从业比例却始终维持在低位,且有进一步走低的趋势。分析这一现象的成因,一方面,物理学专业由于其理论体系复杂、理论知识

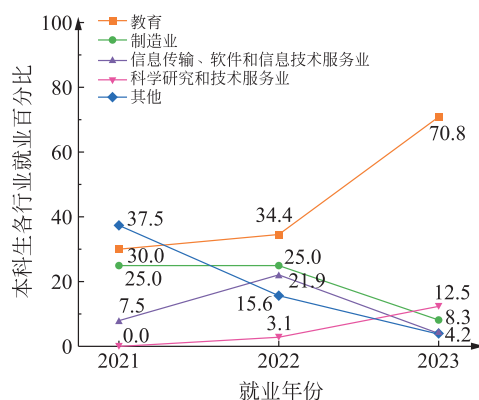


图 9 N 校本科生各行业就业统计

深奥,相较于其他专业往往更需要进一步深造,其传统“对口”岗位为工程师、研究员、高校教师等,这些岗位往往对于应聘者的专业知识和技能水平有较高要求,仅凭本科学历往往难以胜任^[21];另一方面,相较于各类工学专业,物理学作为一门理学课程培养方案往往偏重理论科学而缺少了对于工程实践和产业热点的关注,这就导致物理学毕业生在制造业、信息传输、软件及信息服务业等重视实践能力的行业内与工学毕业生相比缺少竞争力,然而物理相关学科的中小学教师岗位需要较强的理论知识水平和逻辑思维能力,与物理学专业毕业生具有较高的匹配性,且教师行业往往享有较高的社会地位和稳定的收入,因此从事教育行业成了物理学本科毕业生的热门选择^[22]。与此同时,近年来各行业用工需求收缩,就业岗位减少,部分企业存在“学历高消费”的情况,在削减招聘人数、降低薪资待遇的同时又提高了招聘门槛,因为担忧就业和规避竞聘压力,高校毕业生对社会企业就业意愿降低,转而选择教育这一更为稳定的行业。面对日益复杂的宏观经济环境,高校就业指导人员提前帮助学生深入认识经济形势和就业环境,并据此指导学生做出适当规划对学生的职业生涯与未来发展大有裨益^[23]。

7 就业指导策略

7.1 关注学业成绩

在本次物理专业学生毕业去向调查中,我们发现 61% 的学生选择在国内外继续深造。而对于选择直接就业的学生,其职业方向主要集中在物理教育、电子技术、材料科学与物理学紧密相关

的领域。据此可推断:数学、物理学等专业核心知识的掌握,以及英语、计算机等通用技能的熟练运用,对学生的就业升学情况具有显著的影响,而这一观察结果与机器学习算法所揭示的特征重要性相吻合。因此各高校物理类专业应积极关注数学、物理、英语、计算机技术类课程的建设,以增强学生在研究生考试和就业市场中的竞争力。

此外,在机器学习模型中较为重要的课程大多被设置在大一、大二阶段。其中大学英语、线性代数、高等数学、计算机信息技术设置在大一年级,电磁学、电动力学和热力统计学等设置在大二年级。这些课程的成绩可以较为客观地反映学生的学段衔接情况与专业发展情况,将就业指导策略向低年级进行适当调整十分必要。扩大教师就业指导的窗口期,及时关注低年级学生相关课程的成绩,进而对出现明显问题的学生进行早期预警,有助于突破传统就业指导的局限,有针对性地实现对潜在就业风险学生进行干预和预防,从而更好地促进学生的就业和发展。

进一步分析发现,对于物理学专业学生,英语这一文科科目成绩在所有特征中具有最高的重要性。这一现象看似反常,但其反应了当下就业市场的真实需求。在学生从校园走向社会的同时,学生的英语学习的侧重点也从“听”与“读”转移到了“说”和“写”,语言技能的重点由语言的输入转向了输出。理工科学生相较于其他专业学生往往缺少语言表达能力的训练,这就使得具有出色的英语表达能力和写作能力的理工科人才更为稀缺,这些学生在就业市场中往往能够获得更多的优势,在考研与国外升学的竞争中更是如此^[24]。

线性代数、高等数学等数学科目成绩在模型中也具有较高的特征重要性,线性代数、高等数学是理工科学生知识体系的基础,对后续如力学、量子力学、电动力学等重要专业课程均有重要影响。其在工程实践与材料科学领域等就业都有广泛的应用,研究表明线性代数等数学学科与学生就业情况高度相关,是影响学生就业与升学的关键学科^[25]。

电动力学相较于其他物理专业课程在模型预测中具有更高的特征重要性。电动力学是物理专业核心课程之一,学生普遍认为电动力学是四大力学课程中最难掌握的一门学科,其内容涉及电磁场及其性质,学习过程中对于学生的物理基础

与数学能力均有较高要求,这就导致电动力学成绩可以间接反映学生整体专业知识水平^[26]。同时电动力学内容与电力、半导体、芯片、超导材料等热门产业具有较强的关联性,因此电动力学课程应该成为高校就业指导教师关注的重点。

7.2 弱化非学业因素

机器学习数据挖掘的结果表明生源地、性别、民族、政治面貌、农村或城市户口等非学业因素对于学生毕业去向的影响相对较小,决定学生就业升学情况的最主要因素仍为专业知识和通用能力。因此在就业指导过程中,教师应摒弃由非学业因素造成的先入为主的偏见,重点关注学生学业成绩及实践能力等较为客观的因素。在对存在就业困难的个体进行进一步分析时,发现有来自教育欠发达地区农村的同学更有可能存在就业问题。对于这些弱势的同学,校方应给予重视,切实了解学生困难,有针对性地提供帮助与支持,避免学生因经济状况问题放弃深造与就业的机会。

7.3 促进科教融汇、推动产教融合

党的二十大报告中指出教育强国、科技强国、人才强国具有内在一致性,切实推进科教融汇、产教结合是提高人才核心竞争力,强化现代化人才支撑,实现科教兴国战略的重要一环。本研究中N校物理学院本科生毕业去向落实率(即顺利就业或升学的人数占总人数的比例)为94.7%,显著高于《2023中国本科生就业报告》中其他高校同专业88.5%的平均水平,这与该校积极开展实训课程、注重产学结合等措施紧密相关。同时,预警模型中学生实习成绩这一特征具有0.0213的较高重要性指数也很好地印证了这一结论。同时校积极参与各届中国大学生物理学术竞赛、“格致杯”物理师范生教学技能竞赛、全国大学生物理实验竞赛等高水平物理学专业大学生竞赛,该校全专业具有竞赛经验的学生占比9.72%,且这些学生的毕业去向落实率为100%,可见专业知识技能竞赛对于深化理论知识、培养专业技能、提高就业竞争力具有积极影响。近年来由于疫情和国际局势影响,制造业、通讯业等与物理学科高度相关的行业呈现出高度动态和竞争激烈的特点,对人才的要求也日趋多元化和专业化。面对当今就业市场,教育部门与产业界深度合作显得尤为必要。通过建立校企联合培养模式、建立实训基地、开展实践性课程、参与专业知识技能竞赛等多种形式,将实

际产业需求、工作方法及前沿技术引入教育过程中,让学生提前了解行业实际、发展创新能力,将大大利于高校应对就业市场压力,向社会输送更多高素质专业人才。

8 结语

机器学习算法在教育领域具有广泛的应用前景。自2018年教育部发布《教育信息化2.0行动计划》以来,各高校纷纷开启信息化进程,庞大的教育数据资源为机器学习算法提供了丰富的训练样本,有效地解决了机器学习算法在实际应用中面临的数据不足问题。借助强大的算力和卓越的预测效果,机器学习算法可以替代传统教育统计中繁琐的重复性劳动,为教育教学工作者提供更可靠的预测性信息,将传统实证研究与人工智能数据挖掘相结合,深化教育的智能变革,可以为诸多教育教学工作带来新思路和新视角。本研究所采用的三个机器学习算法(随机森林、支持向量机、K近邻)在分析和预测物理专业高校毕业生就业相关问题方面呈现较高($>80\%$)准确率,初现数据挖掘在教育领域的应用价值前景。随着数据量的不断增长,模型的预测效果将进一步提升。将机器学习算法应用于毕业生就业指导,有助于高校实现资源的精准、有效分配,以应对日益严峻的就业形势。

参考文献

- [1] 袁臻. 后疫情时代高校就业指导的困境与对策探析[J]. 中国大学生就业, 2022, (15): 37-42.
YUAN C. Analysis of the dilemma and countermeasures of college employment guidance in the post-epidemic era[J]. China University Students Career Guide, 2022, (15): 37-42. (in Chinese)
- [2] 杨文杰. 高校大学生就业指导困境以及破解策略[J]. 现代职业教育, 2022, (35): 153-155.
YANG W J. Difficulties and solutions in employment guidance for college students[J]. Modern Vocational Education, 2022, (35): 153-155. (in Chinese)
- [3] 杨一平. 大学生就业形势变化与高校就业指导模式的变革[J]. 高等教育研究, 2002(5): 61-63.
YANG Y P. Changes in the employment situation of college students and the transformation of employment guidance models in universities[J]. Journal of Higher Education, 2002(5): 61-63. (in Chinese)
- [4] 宫千千. 新形势下职业生涯规划在大学生就业指导工作中的有效应用[J]. 湖北开放职业学院学报, 2022, 35(2): 46-47, 52.
GONG Q Q. Effective application of career planning in college students' employment guidance under the new situation[J]. Journal of Hubei Open Vocational College, 2022, 35(2): 46-47, 52. (in Chinese)
- [5] 颜中玉, 罗元. 高校“职业发展与就业指导”课面临的困境与改革[J]. 湖南师范大学教育科学学报, 2012, 11(1): 106-109.
YAN Z Y, LUO Y. The challenges and reforms faced by the course of "Career Development and Employment Guidance" in colleges and universities[J]. Journal of Educational Science of Hunan Normal University, 2012, 11(1): 106-109. (in Chinese)
- [6] 赵波, 吴庆畅, 高联雄等. 学生就业贝叶斯网模型的构建与推理[J]. 云南民族大学学报(自然科学版), 2008, (4): 358-361.
ZHAO B, WU Q C, GAO L X, et al. Construction and reasoning of Bayesian network model for student employment[J]. Journal of Yunnan University for Nationalities (Natural Science Edition), 2008, (4): 358-361. (in Chinese)
- [7] 徐岩, 石丽坤, 王明蕙. 基于支持向量机的大学生毕业选择预测研究[J]. 石家庄学院学报, 2008(6): 34-38.
XU Y, SHI L K, WANG M J. Research on predicting graduation choices of college students based on support vector machines[J]. Journal of Shijiazhuang University, 2008(6): 34-38. (in Chinese)
- [8] 何玮. 聚类算法在高校毕业生就业指导中的应用[D]. 广东: 中山大学, 2008.
HE W. Application of clustering algorithm in employment guidance for college graduates[D]. Guangdong: Sun Yat-sen University, 2008. (in Chinese)
- [9] 孙怡帆, 潘昆峰, 孙正阳等. 大学生毕业去向预测的思路与方法——基于机器学习算法的尝试[J]. 教育学术月刊, 2019, (1): 25-35.
SUN Y F, PAN K F, SUN Z Y, et al. Ideas and methods for predicting the graduation destinations of college students: an attempt based on machine learning algorithms[J]. Education Research Monthly, 2019, (1): 25-35. (in Chinese)
- [10] 刘中国, 蒋浩杰, 丁国勇. 大学生学业表现对于就业去向的影响研究——基于多种 EDM 方法[J]. 计算机时代, 2023, (4): 86-90.
LIU Z G, JIANG H J, DING G Y. A study on the impact of college students' academic performance on employment trends: Based on multiple EDM methods[J]. Computer Era, 2023, (4): 86-90. (in Chinese)
- [11] ROY K S, ROOPKANTH K, TEJA V U, et al. Student career prediction using advanced machine learning techniques[J]. International Journal of Engineering & Technology, 2018, 7(2.20): 26-29.
- [12] VIDYASHREERAM N, MUTHUKUMARAVEL A. Student career prediction using machine learning approach

- ches[C]. First International Conference on Computing, Communication and Control System, 2021: 444.
- [13] 黄红娟,徐金荣. 探讨考研热对本科教学的负面影响及对策分析——对应用物理专业考研现状的调查[J]. 人力资源管理(学术版),2009(12):141-142.
HUANG H J, XU J R. Exploring the negative impact of the postgraduate entrance examination fever on undergraduate teaching and analyzing countermeasures: a survey on the current situation of postgraduate entrance examination in applied physics major[J]. Human Resource Management (Academic Edition), 2009(12): 141-142. (in Chinese)
- [14] 王彦辉,祁翠琴,张树海. 协调就业与学业的关系促进高职教育健康发展[J]. 教育与职业,2007(26):179-180.
WANG Y H, QI C Q, ZHANG S H. Coordinate the relationship between employment and academic performance to promote the healthy development of higher vocational education[J]. Education and Vocation, 2007(26): 179-180. (in Chinese)
- [15] 周福盛,黄一帆. 人工智能时代教学中技术理性与教学情感的分离与融合[J]. 课程. 教材. 教法,2023,43(10):50-58.
ZHOU F S, HUANG Y F. The separation and integration of technological rationality and teaching emotions in teaching in the era of artificial intelligence[J]. Curriculum, Teaching Material and Method, 2023, 43(10): 50-58. (in Chinese)
- [16] 黄华,张睿,潘鑫. AI+环境下学生科学探究能力测评研究[J]. 物理与工程,2022,32(6):49-56,65.
HUANG H, ZHANG R, PAN X. Research on the assessment of students' scientific exploration ability in AI+environment[J]. Physics and Engineering, 2022, 32(6): 49-56, 65. (in Chinese)
- [17] 王兴宇. 智能技术在精准教学中的应用及其规制[J]. 高等教育研究,2022,43(4):87-97.
WANG X Y. Research on the evaluation of students' scientific exploration ability in AI+environment: Application and regulation of intelligent technology in precision teaching[J]. Journal of Higher Education, 2022, 43(4): 87-97. (in Chinese)
- [18] 陈良臣,傅德印. 面向小样本数据的机器学习方法研究综述[J]. 计算机工程,2022,48(11):1-13.
CHEN L C, FU B Y. A review of machine learning methods for small sample data[J]. Computer Engineering, 2022, 48(11): 1-13. (in Chinese)
- [19] 宋传颖,易佑斌,李爱民. 生源地视角下大学生就业期望特征分析[J]. 中国大学生就业, 2014,(14):8-12.
SONG C Y, YI Y B, LI A M. Analysis of employment expectation characteristics of college students from the perspective of source areas[J]. China University Students Career Guide, 2014, (14): 8-12. (in Chinese)
- [20] 国家统计局. 2023年中国统计年鉴[M]. 北京:中国统计出版社,2023:8-15.
- [21] 高凯歌. 物理学专业学生就业探讨[J]. 西部素质教育, 2022,8(24):174-177.
GAO K G. Exploration of employment for physics majors[J]. Western China Quality Education, 2022, 8(24): 174-177. (in Chinese)
- [22] 古金霞. “就业窗口”引导地方工院校应用物理学专业人才培养模式探讨[J]. 教育教学论坛,2019,(13):29-31.
GU J X. Exploration of the talent training model for applied physics majors in local engineering colleges guided by the “Employment Window”[J]. Education and Teaching Forum, 2019, (13): 29-31. (in Chinese)
- [23] 张涛,刘全振. 高校毕业生就业工作应急模式创新研究[J]. 江苏高教,2023,(12):101-105.
ZHANG T, LIU Q Z. Research on innovation of emergency mode for employment of college graduates[J]. Jiangsu Higher Education, 2023, (12): 101-105. (in Chinese)
- [24] 温忠义. “英语语言表达能力”概念界定及评价指标研究[J]. 重庆大学学报(社会科学版),2015,21(2):155-161.
WEN Z Y. Research on the definition and evaluation indicators of “English Language Expression Ability” concept[J]. Journal of Chongqing University (Social Science Edition), 2015, 21(2): 155-161. (in Chinese)
- [25] 张雪霞,林升明,马海强. 理工科大学生数学成绩影响因素的统计分析[J]. 西南民族大学学报(自然科学版),2013, 39(1):21-25.
ZHANG X X, LIN S M, MA H Q. Statistical analysis of factors influencing mathematics grades of science and engineering college students[J]. Journal of Southwest Minzu University(Natural Science Edition), 2013, 39(1): 21-25. (in Chinese)
- [26] 谢双媛,王祖源,倪忠强. 信息技术融入电动力学课程建设的思考与实践[J]. 大学物理,2022,41(3):1-8.
XIE S Y, WANG Z Y, NI Z Q. Reflection and practice on integrating information technology into the construction of electrodynamics course[J]. College Physics, 2022, 41(3): 1-8. (in Chinese)