

许惠惠. 基于多分支数据挖掘模型的大学生精准思政研究[J]. 智能计算机与应用, 2025, 15(4): 115-120. DOI: 10.20169/j.issn.2095-2163.24062702

基于多分支数据挖掘模型的大学生精准思政研究

许惠惠

(山西药科职业学院 素质教育教学研究部, 太原 030031)

摘要: 本研究旨在利用数据挖掘的方法, 构建高职院校不同年级学生的群体特征画像, 以开展定制化的精准思想政治教育研究。本文运用数据挖掘模型, 收集学生基本信息、人格特质以及思想政治课程情况等多方面数据, 绘制高职院校学生的群体特征轮廓。这有助于更深入地了解学生在思想政治和人格特质方面的特点、需求与倾向。同时, 将研究学生思想政治情况与人格特征之间的相关性。通过对学生群体特征的深入分析, 研究结果可为高职院校制定更加精准的思想政治教育策略提供依据。

关键词: 数据挖掘模型; 学生画像; 精准思政

中图分类号: G420

文献标志码: A

文章编号: 2095-2163(2025)04-0115-06

Research on ideological and political education in vocational colleges based on data mining model

XU Huihui

(Department of Quality Education Teaching and Research, Shanxi Pharmaceutical Vocational College, Taiyuan 030031, China)

Abstract: This study aims to utilize data-driven approaches to construct a demographic profile of students across different grades in vocational colleges, facilitating customized research in political and ideological education. Specifically, this paper applies data mining models to gather various data points including students' basic information, personality traits, and participation in ideological and political courses, to delineate the demographic profile of vocational college students. This research contributes to a deeper understanding of students' characteristics, needs, and inclinations in ideological and political aspects as well as personality traits. Moreover, the paper investigates the correlation between students' ideological and political situations and their personality traits. Through in-depth analysis of student demographic characteristics, the research findings could serve as a basis for vocational colleges to formulate more precise and tailored strategies for ideological and political education.

Key words: data mining model; student profile; precision ideological and political education

0 引言

精准思政是一种基于前沿技术、特别是大数据的思政工作模式。该技术是以思维为指导, 以精准化的方式来展开思想政治教育活动^[1]。该模式的核心在于通过对受教育者多个维度的数据信息进行分析, 以制作标签的方式描绘受教育者的用户画像, 并通过这些画像来有针对性地设计教育策略。这种方法强调个性化和智能化教育, 通过深入了解学生

的特点和需求, 以及借助先进的技术手段, 实现教育的精准化和智能化^[2]。这一模式的目标是提高思想政治教育的效果和影响力, 以更好地满足学生的需求, 其中从数据驱动到实现智能教育的推进路线如图1所示。

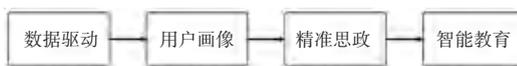


图1 数据驱动到智能教育的路线图

Fig. 1 A roadmap for data-driven to smart education

基金项目: 2024年度全国高等职业院校信息技术课程教学改革研究项目(KT2024176); 2021年度山西省高等学校哲学社会科学研究项目(思想政治教育专项)(2021zsszsx207); 山西省高职院校思想政治教育研究会2021年度思想政治教育研究项目(SYH2021-032)。

作者简介: 许惠惠(1983—), 女, 硕士, 讲师, 主要研究方向: 计算机应用, 数据分析与挖掘。Email: 95201157@163.com。

收稿日期: 2024-06-27

在精准思政中,数据驱动是一个重要的起点。借助大数据和其他前沿技术,教育者可以收集、分析和利用多维度的数据信息。这些数据可以包括学生的学习成绩、社交网络活动、心理健康状况、兴趣爱好等。通过对这些数据的深入分析和挖掘,可以更好地了解学生的特点和需求^[3]。

在精准思政中,构建学生的用户画像是一个关键的步骤。用户画像是对学生进行细致刻画的过程,通过对学生的个性、兴趣、价值观等方面进行综合分析,形成一个全面而准确的描述^[4]。用户画像可以帮助教育者更好地了解学生的特点和需求,从而制定有针对性的教育策略。

基于学生的用户画像,精准思政可以设计个性化的教育策略。个性化教育策略是根据学生的特点和需求,量身定制的教育方案。通过精准思政,教育者可以根据学生的兴趣、学习风格、认知水平等因素,提供符合其个性化需求的教育内容和方法,从而提高教育的针对性和有效性^[5]。

最后,精准思政实现了教育的智能化。借助大数据和人工智能等技术手段,精准思政为教育者提供智能化的决策支持和教学辅助。通过对大量数据的分析和挖掘,结合智能算法和模型,精准思政可以帮助教育者更好地了解学生的学习状态和需求,提供个性化的学习推荐和反馈,促进学生的全面发展^[6]。

学生画像和精准思政之间存在密切的关系。通过对学生画像的分析,可以了解学生的思想倾向、关注点和问题所在,从而更好地制定和实施精准思政教育的策略和方法。学生画像可以帮助教师和教育

管理者更深入地了解学生的需求和特点,针对不同学生的个性化差异,有针对性地进行教育引导和思政教育。通过学生画像的分析,可以为学生提供个性化的精准思政教育,使其在参与中更容易接受和理解思政教育的内容。同时,通过了解学生的学习情况和兴趣爱好,可以将思政教育与其他学科相结合,使思政教育更加贴近学生的实际需求,提高思政教育的针对性和有效性。

本文学生画像的构建从学生的基本信息、人格特质和思政课程三个方面入手。首先,学生的基本信息包括姓名、年龄、性别以及所在学校和年级等。这些信息能够提供学生的身份背景、成长环境和社交圈等方面的线索,为了解学生提供基础。其次,人格特质对于了解学生的行为和学习方式至关重要。卡特尔人格因素量表和 Big Five 人格量表^[7]是常用的评估工具。通过评估学生在外向性、神经质、开放性、宜人性和尽责性等5个维度上的特征,可以描绘出学生的个性特点。最后,思政课程是培养学生思想道德素质和法治意识的重要环节。通过观察学生在思政课程中的表现和成绩,可以了解其对中国特色社会主义理论与实践的掌握程度和理解深度,这包括政治觉悟、道德品质、法治意识和社会责任感等方面。

1 数据采集

数据采集是构建学生画像的关键步骤,涵盖了学生的基本信息、人格特质和思政课程等方面,详情如图2所示。为了获取这些数据,对不同的数据类别采用不同的收集方法。

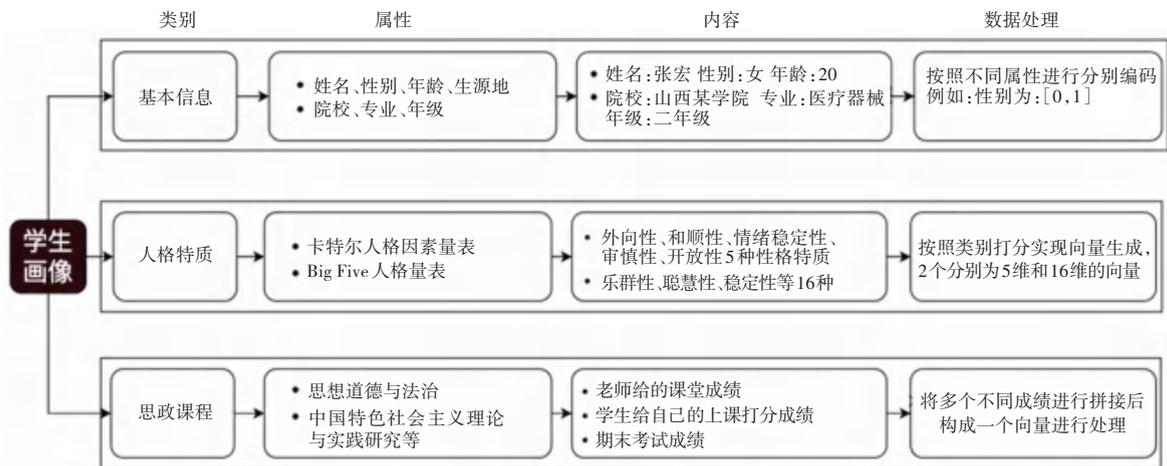


图2 学生画像数据来源和属性内容

Fig. 2 Student portrait data sources and attribute content

首先,通过学校档案对学生的基本信息进行收集。学校在学生入学时通常会记录学生姓名、年龄、性别以及所在学校和年级等基本信息。这些数据可以提供学生的身份背景和学习环境,为学生画像提供基础,从而为构建学生的背景信息和学生的各维度画像提供了可能。

其次,通过使用卡特尔人格因素量表和 Big Five 人格量表对学生的人格特质进行评估。这些量表中包含一系列问题和评价项,学生可以根据自身特点做出回答或进行评分。通过分析和解读学生的回答和评分,可以了解学生的外向性、神经质、开放性、宜人性和尽责性等人格特质。其中, Big Five 人格量表包括以下人格属性。

(1)外向性:部分学生表现出较高的外向性,愿意与他人交流和合作,展现较强的社交能力。

(2)神经质:学生呈现出较高的神经质,可能更容易焦虑和紧张,需要更多的支持和关怀。

(3)开放性:大多数学生展现出较高的开放性,对新事物和想法持开放态度,具有较强的求知欲和创造力。

(4)宜人性:学生普遍表现出较高的宜人性,善良友善、乐于助人,有着良好的人际关系。

(5)尽责性:部分学生表现出较高的尽责性,责任心强,对任务的完成有较高的自我要求和执行力。

最后,思政课程方面的数据通过多种途径进行收集,包括教师观察学生在课堂上的表现、参与度和互动情况来评定学生的课堂成绩,同时收集学生对自己的上课打分成绩,以及学生的期末考试成绩。

2 数据挖掘模型

为了实现对学生的画像的刻画和对学生进行精确思政评价,本论文构建了三分支模型,包含了逻辑回归、随机森林和支持向量机三个分支,最后通过多层感知器层将其输出拼接在一起^[8]。首先是逻辑回归分支,以原始特征数据为输入,并直接连接到特征输出层,输出层使用逻辑回归模型将输入特征映射到二分类的概率。随机森林分支也接受原始特征数据作为输入,通过决策树的集成学习方式,进行多个决策树的投票,最终得到分类结果。支持向量机分支则通过寻找一个超平面来将数据分隔开,使得不同类别的样本尽可能远离超平面。最后,多层感知器层将3个分支的输出进行拼接,形成最终的预测

结果。这种结构能够充分利用3种不同模型的特点,提高整体的分类性能。

本文的主要数据挖掘模型如图3所示,以下将主要介绍研究模型中各个模块的设计和构建。

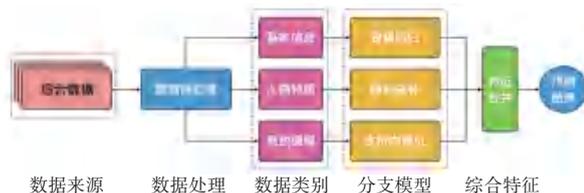


图3 模型结构图

Fig. 3 Model structure

3 模型结构介绍

3.1 基于逻辑回归的特征提取

逻辑回归是一种用于解决分类问题的统计学习方法。尽管名称中带有“回归”,但逻辑回归实际上用于处理分类任务,而不是回归任务,其主要思想是通过将输入特征的线性组合应用逻辑函数,将数据映射到一个二元的输出^[9-10]。基本信息中的年龄和专业在一定程度上对学生的人格特征会有影响作用,基于此考虑,通过使用逻辑回归并行地提取处理学生数据中的6个主要背景因素,将相关指标数据作为输入,从而提取出与学生综合分数高度相关的特征表征。

逻辑回归通过将线性函数的输出应用于逻辑函数(Sigmoid函数)来进行分类。假设有一个输入特征向量 $x = \{x_1, x_2, \dots, x_n\}$, 则逻辑回归模型可以表示为:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (1)$$

其中, z 表示线性函数的输出, $\{\beta_0, \beta_1, \beta_2, \dots, \beta_n\}$ 是模型的参数。然后,将 z 应用于逻辑函数 $\sigma(z)$, 得到分类概率 $P(Y = 1 | x)$:

$$p(Y = 1 | x) = \sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

其中, Y 表示分类结果。

3.2 基于随机森林的特征提取

随机森林是由多个决策树组成的集成模型。每棵决策树通过对随机抽取的样本和特征进行训练。对于分类问题,随机森林中的每棵树都会输出一个类别,并通过投票或取平均值的方式来确定最终的预测结果^[11-12]。

3.3 基于支持向量机的特征提取

SVM的基本形式是线性分类器^[13-15],其决策边界可以表示为:

$$W \cdot x + b = 0 \quad (3)$$

其中, W 表示法向量, b 表示偏置项。对于二元分类问题, 决策函数为:

$$f(x) = W \cdot x + b \quad (4)$$

对于训练数据中的每个样本 (x_i, y_i) , 这里 y_i 表示类别标签, SVM 的目标是找到最大间隔超平面, 使得所有样本点到该超平面的距离最大化, 并满足以下约束条件:

$$y_i(W \cdot x + b) \geq 1, i = 1, 2, \dots, n$$

其中, n 表示训练样本的数量。

3.4 综合特征合并

在使用逻辑回归、随机森林和支持向量机对学生的基本信息、人格特质和思政课程进行各类特征提取后, 研究将通过多特征拼接操作实现对学生综合特征的提取。在获得综合特征后, 使用多层感知器 (Multilayer Perceptron, MLP) 实现对学生的不同类型的映射。

4 实验

4.1 实验参数和环境设置

在本实验中, 采用了一台配备 NVIDIA GeForce RTX 3090 GPU 的工作站进行深度学习模型的训练和评估。研究使用 PyTorch 深度学习框架构建和训练了本文提出的模型。训练过程使用了批量梯度下降 (Batch Gradient Descent) 优化算法^[16], 学习率设置为 0.001, 并采用了交叉熵损失函数作为模型的损失函数。

4.2 性能评估指标

研究中将数据集划分为训练集、验证集和测试集, 其中训练集占 70%, 验证集占 15%, 测试集占 15%。在训练过程中, 使用了早停法 (Early Stopping) 来防止模型过拟合^[17-20], 并在验证集上进行了模型性能的监控。

研究中评估了模型在测试集上的性能, 主要包括准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1 分数。这些指标是评估分类模型性能的常用指标。其中, Accuracy 衡量了模型正确分类的样本比例; Precision 衡量了模型在预测为正类别的样本中的正确率; Recall 衡量了模型在所有正类别样本中被正确预测的比例; F1 分数是精确率和召回率的调和平均值, 综合考虑了模型的准确性和全面性。

4.3 实验结果分析

本文提出的复合模型与传统机器学习方法 (如

逻辑回归、支持向量机等) 进行了性能对比。结果显示, 本文提出的深度学习模型在准确率、精确率、召回率和 F1 分数上均显著优于对比方法。具体对比结果见表 1。

表 1 数据挖掘方法实验结果

Table 1 Experimental results of data mining methods

模型	准确率	精确率	召回率	F1 分数
本文模型	0.85	0.87	0.83	0.85
逻辑回归	0.78	0.75	0.80	0.77
支持向量机	0.80	0.79	0.78	0.79
K 近邻	0.75	0.72	0.77	0.74
朴素贝叶斯	0.72	0.70	0.75	0.72

从表 1 可以看出, 本文的模型在所有指标上均取得了最佳性能, 证明了其在该任务上的有效性和优越性。

4.4 收敛情况对比

在模型训练过程中, 研究监测了本文的模型在训练集上的损失变化情况, 具体如图 4 所示。损失函数的变化反映了模型在训练过程中的收敛情况和拟合能力。从损失变化曲线可以观察到, 随着训练轮次的增加, 训练集上的损失都呈现下降趋势, 并逐渐收敛到一个较低的水平。

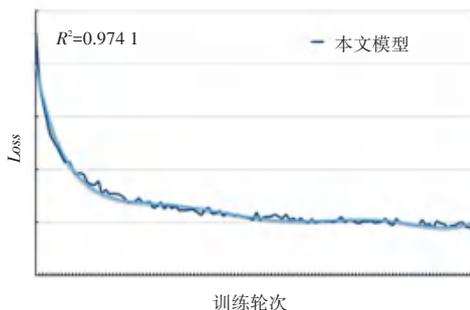


图 4 模型训练收敛情况

Fig. 4 Convergence of model training

4.5 案例分析

在分析学生在人格特质和思政课程方面的表现时, 研究着重对不同的人格特质的思政情况进行了精确分析, 不同分数学生和人格属性之间的雷达图如图 5 所示, 对应的具体分数见表 2。以下是从图 5 中得出的分析结果:

(1) 外向性。从结果中可知, 外向性高的学生倾向于积极参与思政课程中的讨论和活动, 并在课堂上表现出较高的主动性和社交能力。同时也更愿意与同学分享自己的观点和经验, 对于团队合作和集体活动也表现出较强的积极性。

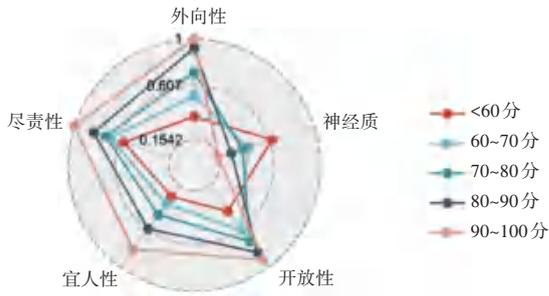


图5 不同分数学生和 Big Five 人格属性之间的雷达图

Fig. 5 Radar graph between students with different scores and Big Five personality attributes

(2)神经质。神经质较高的学生可能对于思政课程中的一些话题或观点产生较强的情绪波动,可能更容易受到负面情绪的影响。因此,在思政课程教学中,需要对这部分学生的情感状态给予更多关注,采取措施助其保持情绪稳定,并引导这部分学生

正确理解和面对社会现实。

(3)开放性。开放性高的学生对于新鲜事物和不同观点有较强的接受和包容性,在思政课程中可能更愿意探讨一些较为前沿和复杂的议题。因此,思政课程可以通过引导这部分学生参与讨论和研究,促进其个人成长和思维发展。

(4)宜人性。宜人性高的学生通常具有较强的合作精神和同理心,在思政课程中更容易与同学和教师建立良好的互动关系。这部分学生乐于助人,能够在团队中发挥积极的作用,促进集体的发展和进步。

(5)尽责性。尽责性高的学生具有较强的自律能力和责任感,在思政课程学习中表现出较强的执行力和毅力。这部分学生对于学习任务和社会责任都能够认真对待,乐于承担起自己的责任,为社会和国家的发展做出积极的贡献。

表2 不同等级学生的 Big Five 人格属性得分

Table 2 Big Five personality attribute scores

思政课程分数/分	外向性	神经质	开放性	宜人性	尽责性
<60	0.347 8	0.628 3	0.419 4	0.267 1	0.562 3
60~70	0.521 4	0.413 7	0.621 9	0.342 9	0.679 5
70~80	0.718 2	0.367 4	0.735 6	0.456 2	0.723 8
80~90	0.923 7	0.267 8	0.847 3	0.602 1	0.829 1
90~100	1.000 0	0.154 2	0.924 6	0.817 5	1.000 0

5 结束语

本研究深入研究学生的基本信息、人格特质和思政课程,对不同学生的人格特质和思政情况进行精确分析,研究构建机器学习模型,利用数据挖掘方法,发现了不同类别学生的特点。

成绩较低的学生可能具有一定的敏锐性和变革性,但缺乏稳定性和自律性;成绩在80~90分的学生展现出优秀的自律性和规范性,但可能对变革和创新持保守态度;成绩达到90~100分的学生通常表现出卓越的稳定性和自律性,但可能过于自信和独立。

这些发现为高职院校思政课程的教学和管理提供了重要参考,有助于制定更精准的教学策略和个性化辅导方案,促进学生成长成才。

参考文献

[1] 刘任佳. 基于大学生画像分析的精准思政探究[J]. 上海第二工业大学学报, 2023, 40(4):385-391.

[2] 葛丽亚. 大数据赋能综合评价为学生成长精准画像[J]. 河北教育(德育版), 2024, 62(2):14-15.

[3] 杨怡雪,齐明明. 基于学生画像的高校思想政治理论课精准教学研究[J]. 沈阳工程学院学报(社会科学版), 2024, 20(1):76-82.

[4] 李欣,刘春,贾美娟,等. 面向计算机公共课的思政资源平台设计与开发[J]. 智能计算机与应用, 2023, 13(2):86-91.

[5] 梁肇敏,梁婷婷. 基于大数据技术的电商用户画像可视化系统设计及实现[J]. 智能计算机与应用, 2023, 13(3):242-245.

[6] 骆睿鹏,邹任玲,孟令鹏,等. 基于机器学习的抑郁症脑电识别研究[J]. 智能计算机与应用, 2023, 13(11):82-87.

[7] 罗杰,戴晓阳. 中文形容词大五人格量表的初步编制 II:测验效度[J]. 中国临床心理学杂志, 2015, 23(4):571-575.

[8] 陈添源,梅鑫. 多源数据融合的用户画像识别与推荐实证研究[J]. 情报理论与实践, 2024, 47(4):171-180.

[9] 申冰可,邵林芳. 决策树模型与 logistic 回归探究大学生入党意愿自身影响因素及结果预测[J]. 电脑知识与技术, 2022, 18(8):26-28.

[10] MIYAZAKI Y, KAWAKAMI M, KONDO K, et al. Logistic regression analysis and machine learning for predicting post-stroke gait independence: a retrospective study[J]. Scientific Reports, 2024, 14(1):21273-21273.

[11] HU Jianchang, SZYMCZAK S. A review on longitudinal data analysis with random forest[J]. Briefings in Bioinformatics,

- 2023, 24(2): bbad002.
- [12] SUN Zhigang, WANG Guotao, LI Pengfei, et al. An improved random forest based on the classification accuracy and correlation measurement of decision trees [J]. *Expert Systems with Applications*, 2024, 237: 121549.
- [13] 朱范炳, 陈泽, 张翔. 基于改进蜂群算法优化的支持向量机研究与应用[J]. *智能计算机与应用*, 2023, 13(8): 197-200.
- [14] LIANG Hexin, WANG Zongying, LI Yao, et al. The application value of support vector machine model based on multimodal MRI in predicting IDH-1 mutation and Ki-67 expression in glioma[J]. *BMC Medical Imaging*, 2024, 24(1): 244.
- [15] WANG Huajun, SHAO Yuanhai. Fast generalized ramp loss support vector machine for pattern classification [J]. *Pattern Recognition*, 2024, 146: 109987.
- [16] CHEN Liuhong, XIONG Meixin, MING Ju, et al. Efficient mini-batch stochastic gradient descent with Centroidal Voronoi Tessellation for PDE-constrained optimization under uncertainty [J]. *Physica D: Nonlinear Phenomena*, 2024, 467: 134216.
- [17] EDWARDS L B, JACKSON S E, KEFALIANOS E, et al. Contemporary clinical conversations about stuttering: Can intervention stop early stuttering development? [J]. *International Journal of Speech-language Pathology*, 2024(1): 1-6.
- [18] AL-RIMY B A S, SAEED F, AL-SAREM M, et al. An adaptive early stopping technique for DenseNet169-based knee osteoarthritis detection model [J]. *Diagnostics*, 2023, 13(11): 1903.
- [19] KASAI M, MURATA H, NAKAHIRA T, et al. Effect of early stopping on error performance of iterative MIMO equalization: An experience in reality [J]. *IEICE Communications Express*, 2020, 9(10): 489-494.
- [20] BEEHARRY Y, FOWDUR P T, SOYJAUDAH S M K. Performance of bit level decoding algorithms for binary LTE Turbo codes with early stopping [J]. *Journal of Electrical & Electronics Engineering*, 2017(2): 3399-3415.