

基于K-Medoids算法的大学生体质分类评价模型研究

刘禹宏¹, 张琛煜¹, 仇周亮², 耿秀丽^{1*}

¹上海理工大学管理学院, 上海

²上海理工大学体育教学部, 上海

收稿日期: 2024年4月18日; 录用日期: 2024年6月22日; 发布日期: 2024年6月30日

摘要

大学生是国家发展重要的后备力量。然而缺乏科学规律的身体锻炼以及熬夜等生活习惯问题使得大学生体质健康检测结果不理想。为了深入分析大学生体质情况现状, 建立了一种基于K-medoids算法的大学生体质健康分类评价模型。以某高校的实际体测数据为样本, 进行了案例分析, 所得结果可应用于学生体质改善方案的制定中。

关键词

K-Medoids算法, 大学生体质健康, 分类评价模型

Research on the Classification and Evaluation Model of College Students' Physique Based on K-Medoids Algorithm

Yuhong Liu¹, Chenyu Zhang¹, Zhouliang Qiu², Xiuli Geng^{1*}

¹Business School, University of Shanghai for Science and Technology, Shanghai

²Department of PE, University of Shanghai for Science and Technology, Shanghai

Received: Apr. 18th, 2024; accepted: Jun. 22nd, 2024; published: Jun. 30th, 2024

Abstract

College students are an important reserve force for national development. However, the lack of

*通讯作者。

文章引用: 刘禹宏, 张琛煜, 仇周亮, 耿秀丽. 基于 K-Medoids 算法的大学生体质分类评价模型研究[J]. 运筹与模糊学, 2024, 14(3): 1185-1193. DOI: 10.12677/orf.2024.143349

scientific and regular exercise lifestyle and living habits such as staying up late make college students perform poor in the physical health test of college students. In order to deeply analyze the current situation of college students' system, we establish a classification and evaluation model of college students' physical health based on K-medoids algorithm. The actual physical data of a university is taken as a sample to make a case analysis. The results can be used in the formulation of students' physical fitness improvement program.

Keywords

K-Medoids Algorithm, College Students' Physical Health, Classification Evaluation Model

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

2021 年发布的第八次全国学生体质与健康调研结果显示,我国学生体质健康达标优良率总体呈上升趋势,但大学生体质健康达标优良率基本上没有增长[1]。《国家学生体质健康标准》[2]是测量学生体质健康状况和锻炼效果的评价标准,该标准通过 5 个项目将大学生体质分为优秀、良好、及格和不及格四个等级。但是该标准并不能反映各地区各高校大学生实际的体质健康分类,难以辅助高校制定科学合理的大学生运动指导计划。此外,大学生往往难以客观地进行自我健康评价,根据“丁香医生”与中国青年报联合发布的《2020 中国大学生健康调查报告》,大学生群体对自己的总体健康状态自评偏高且相对于疾病问题,更加关注生活相关健康问题,如皮肤状态,睡眠不足,情绪问题等[3],这使得大学生往往长期处于亚健康状态而不自知,最终错过身体调理与疾病预防的最佳时机。

针对学生体质评价问题,目前主流研究模式大多采用统计学方法。吕宏蓬,王天芳等[4]通过分析 913 例大学生体质,采用 SPSS 25.0 软件分析数据;各体质类型分布情况以频数、百分率描述,所有检验采用双侧检验, $P < 0.05$ 为差异有统计学意义,将 913 例学生分成九种体质类型。安义德等[5]通过对女大学生设计中等强度运动处方,进行干预,使用 SPSS 20.0 进行统计分析,组间身体素质指标的比较采用独立样本 t 检验,从而对受试者体质水平进行划分。《中国人群精准运动处方研制与运动处方库建设》课题组通过对 116 个普通人群精准运动处方和 70 个慢病患者精准运动处方对获得数据进行统计学分析及绘图,进行 Meta 分析,从而依次对不同体质患者进行分类,研究出 130 个普通人群精准运动处方,100 个慢性病患者精准运动处方[6]。结合机器学习算法的应用,本文提出了基于 K-medoids 算法的大学生体质分类评价模型,通过机器学习对大学生体质进行分类,从而获得更加科学精准,更具代表性的学生体质分类结果。

2. 大学生体质分类评价模型构建

以不同体育运动班作为教学单位的传统教学模式一直是我国高校体育工作中的主要模式[7]。然而,在同一运动班中,不同学生对于同一项运动的掌握能力会受先天因素或后天学习因素影响,同时不同年级的学生也会反映出不同的运动水平能力,使学生在运动表现方面具有差异,在传统教学过程中,老师会根据学生的个体运动表现,通过主观判断实施不同的教学方案,本研究通过设计大学生体质分类评价模型,运用大数据分析的方法[8],收集学生运动过程中的客观数据,从而形成更科学、准确的评价模式,

辅助老师进行因材施教的教学工作，针对不同特点人群开展针对性教学，实现差异化、个性化。研究流程图见图 1。

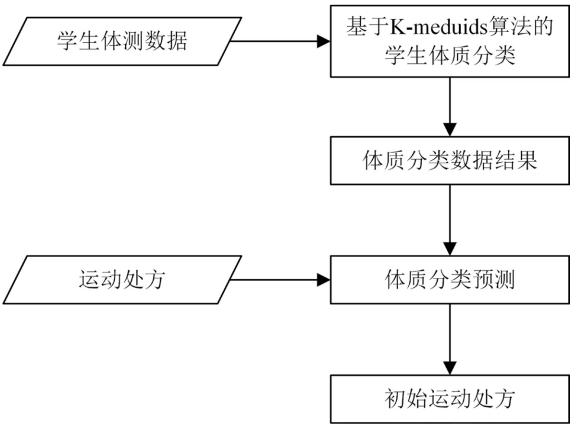


Figure 1. Research flow design
图 1. 研究流程设计图

本研究以《国家学生体质健康标准》为依据，以标准要求的各单项指标包括：BMI，肺活量，50 米跑，立定跳远，坐体前屈，仰卧起坐(女)/引体向上(男)，1000 米跑(男)/800 米跑(女)，为依据，通过收集大量数据，建立数据表格，并剔除缺失值与异常值，如缺少单项成绩，或是如短跑，肺活量为 0 的学生，去除噪声数据。同时，为了在进行数据分析时统一标准，避免不同成绩与评价标准带来的衡量标准异常，我们针对数据进行预处理，通过归一化手段，将数据范围限制到 $[-1, 1]$ ，有助于后续进行降维处理[9]，归一化后的样本数据分布结果如图 2 所示。

K-medoids 算法被称为 k 中心点算法，是一种聚类算法[10] [11]。与常用的 K-means 相比，K-medoids 的取值只能是数据样本范围中的样本，其分类参考点为集群中位于最中心的对象，而不是将集群中的平均值作为参考点，因此能够排除异常值带来的干扰，使聚类结果更加具有准确性，代表性。如图所示，实验数据样本具有中心聚集，散点分布的特点，周围数据容易对最终实验结果造成干扰，因此我们采用 K-medoids 算法，规避噪声数据，使实验结果更加精确。

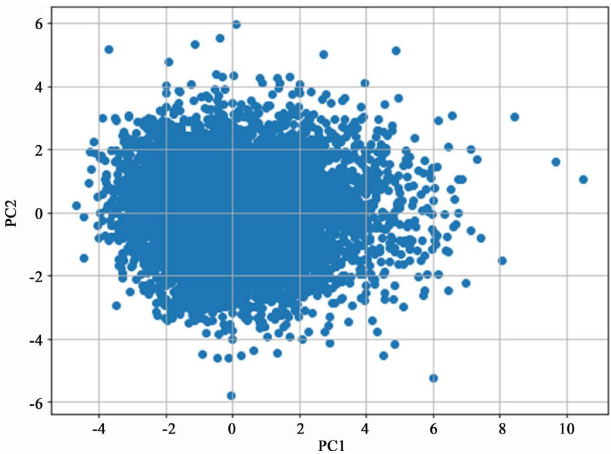


Figure 2. PCA dimensional reduction results of sample data
图 2. 样本数据 PCA 降维结果

所提基于 K-medoids 聚类的大学生体质分类步骤如下。

步骤 1: 通过与学校体育教学部开展合作, 采集上海理工大学 2022 年大一至大四全体学生体测数据, 包括 BMI、肺活量、立定跳远、引体向上、坐位体前屈、50 米跑、100 米跑、引体向上/仰卧起坐共八项数据, 并录入矩阵, 针对数据进行整理筛选。

步骤 2: 根据预处理数据计算各样本点间欧氏距离, 放入矩阵中, 并确定聚类中心点(标准类别)个数 k 。

步骤 3: 随机选取 k 个体测数据作为中心点。

步骤 4: 计算各点与中心点距离, 并分配至与其距离最近的中心点。

步骤 5: 计算各数据点至其集合内其他数据点的距离并重新分配中心点。

步骤 6: 当中心点停止更新迭代时, 则获得最终结果即各标准类别代表性数据。

基于 K-Medoids 算法的学生体质分类算法流程见图 3。

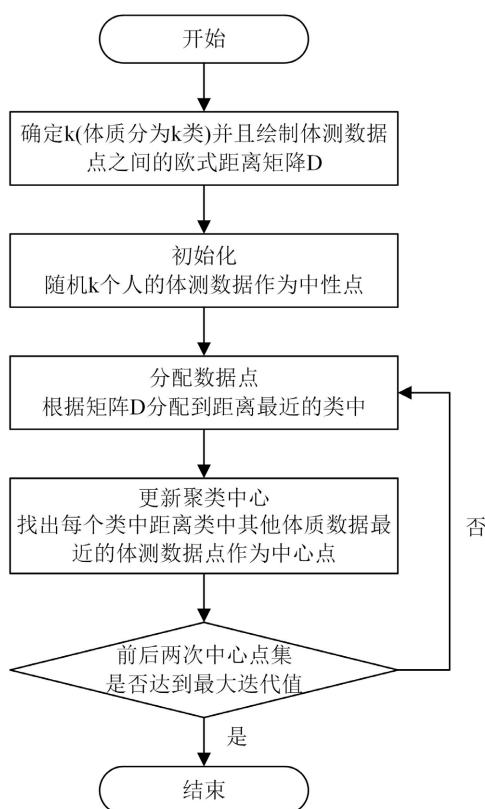


Figure 3. K-medoids algorithm flow chart

图 3. K-medoids 算法流程图

3. 案例分析

3.1. 研究对象

本次研究采用上海理工大学 2022 年大学生体质健康测试数据, 以各年级各体质水平大学生为研究对象。根据性别进行单独研究, 最终经过数据清洗, 共收集男生数据 6656 条, 女生数据 5355 条。

3.2. 数理统计法

通过对数据进行预处理, 将不同体测项目数据归纳至同一范围, 对数据进行标准化处理, 其后采用

Scikit-learn 机器学习库中的 matplotlib 和 sklearn 函数, 对归一化后的数据进行聚类分析, 得到具有代表性的学生体质的数据进行分析, 从而获得具体成绩。

3.3. 数据收集与预处理

由于大学生体测数据量较大, 在收集了 2022 年上海理工大学大学生体测数据后, 需要先对数据进行整理, 去除噪声数据, 去除缺失项及异常值, 并对数据进行归一化整理, 使数据限定至 $[-1, 1]$, 有助于聚类算法的运行。部分随机数据及归一化的数据见表 1、表 2。

Table 1. Male random data

表 1. 男性随机数据

BMI (kg/m^2)	肺活量 (毫升)	立定跳远 (cm)	坐位体前屈 (cm)	50 米 (秒)	引体向上 (个)	1000 米 (分)
26.07	3304	7.7	205	12.2	30	5.36
22.81	3094	6.5	240	22.4	10	3.2
18.28	3950	8	252	21	7	3.36
20.26	4978	7.1	230	11.8	4	4.01
21.77	2924	6	195	2.6	6	5.35
17.06	2828	7.8	225	11.6	2	4.25
22.07	3217	6.9	225	20.6	6	4.13
23.46	2841	8.1	200	8.3	2	5.34
18.01	2899	7.9	217	4	0	4.5
23.24	5263	7.4	225	2	2	4.38

Table 2. Male random data normalization results

表 2. 男性随机数据归一化结果

BMI (kg/m^2)	肺活量 (毫升)	立定跳远 (cm)	坐位体前屈 (cm)	50 米 (秒)	引体向上 (个)	1000 米 (分)
0.3175	0.314812968	0.3275862	0.5	0.4573864	0.41666667	0.605405405
0.2277	0.293865337	0.1206897	0.705882353	0.6022727	0.13888889	0.216216216
0.1028	0.37925187	0.3793103	0.776470588	0.5823864	0.09722222	0.245045045
0.1573	0.48179551	0.22418	0.647058824	0.451705	0.05555556	0.36216216
0.1989	0.27690773	0.034483	0.441176471	0.321023	0.0833333	0.60360361
0.0692	0.26733167	0.344828	0.617647059	0.448864	0.0277778	0.40540541
0.2073	0.30613466	0.189655	0.617647059	0.576705	0.0833333	0.38378378
0.2456	0.26862843	0.396552	0.470588235	0.401987	0.0277778	0.6018018
0.0955	0.27441397	0.362069	0.570588235	0.340909	0	0.45045045
0.2395	0.51022445	0.275862	0.617647059	0.3125	0.0277778	0.42882883

3.4. 数据处理结果

将采集到的数据放入矩阵, 确定聚类中心点数 k , 随机抽取 k 个点作为中心点, 随后使用 K-medoids

算法, 通过计算各点间欧氏距离, 根据距离矩阵将各个样本点分到距离它最近的中心点集中, 重新计算各点集中心, 再次将各点分配至各中心点, 循环迭代, 直至最后一次输出结果与上次相同则停止迭代, 输出结果。

将 6656 条男生与 5355 条女生数据输入程序并使用 K-medoids 聚类分析后, 分析 BMI, 肺活量, 50 米跑, 立定跳远, 坐体前屈, 仰卧起坐(女)/引体向上(男), 1000 米跑(男)/800 米跑(女)。通过与体育部教师商议讨论, 最终确定将 K 值确定为 5, 以满足不同体质水平人群的健康需求, 方便教师开展教学以及制定个性化运动处方[12]。男生聚类中心点结果如表 3 所示。女生聚类中心点结果如表 4 所示。

Table 3. Male clustering center point results
表 3. 男生聚类中心点结果

序号	BMI (kg/m ²)	肺活量 (毫升)	立定跳远 (cm)	坐位体前屈 (cm)	50 米 (秒)	引体向上 (个)	1000 米 (分)
1	0.1798	0.34384	0.20689	0.735294118	0.529829	0.111111	0.36216216
2	0.2053	0.329177	0.29311	0.588235294	0.589488	0.055556	0.4054054
3	0.1536	0.324389	0.24137	0.647058824	0.375	0.069444	0.41441441
4	0.3349	0.355112	0.25862	0.588235294	0.497159	0.027778	0.44684684
5	0.2776	0.350822	0.37931	0.470588235	0.399147	0.013889	0.54054054

Table 4. Female cluster center point results
表 4. 女生聚类中心点结果

序号	BMI (kg/m ²)	肺活量 (毫升)	立定跳远 (cm)	坐位体前屈 (cm)	50 米 (秒)	引体向上 (个)	1000 米 (分)
1	0.199429	0.178843	0.210526	0.59375	0.899329	0.466667	0.186471667
2	0.205488	0.181119	0.276316	0.46875	0.581208	0.522222	0.301645338
3	0.217817	0.174449	0.276316	0.425	0.440269	0.366667	0.283363803
4	0.199967	0.178922	0.407895	0.3125	0.502013	0.422222	0.354661796
5	0.364106	0.196029	0.315789	0.375	0.522148	0.444444	0.360146252

3.5. 聚类结果分析

通过分析聚类结果, 我们发现与教育部标准, 即以 90 分、80 分、60 分对学生体质进行优秀、良好、合格及不合格四类划分相比, 基于 K-medoids 的大学生体质分类模型可根据用户需求进行更加精细的多层次学生体质分类, 同时, 通过运用该模型针对大学生体质展开分析, 我们发现不同体质水平学生, 其运动表现所反映出的分数段更加密集, 不同体质水平学生间差异并不明显显著, 总体呈正太分布态势。男生聚类结果分数见图 4, 女生聚类结果分数见图 5。

如图 6、图 7 关于男女生各项运动水平聚类结果折线图所示, 不同体质水平的大学生其运动表现水平差异总体较为集中, 通过运用基于 K-medoids 的大学生体质分类模型, 重新排列的大学生体质分类更加明显的展示出不同体质水平的间的具体差异以及反映在运动表现中的差异, 由此可以帮助学生及老师在开展日常体育锻炼及教学活动中, 有针对性地开展特定训练, 提高学生特定运动表现水平, 从而针对性提高其对应的诸如心肺能力, 柔韧性等身体素质, 帮助学生提高体质水平。

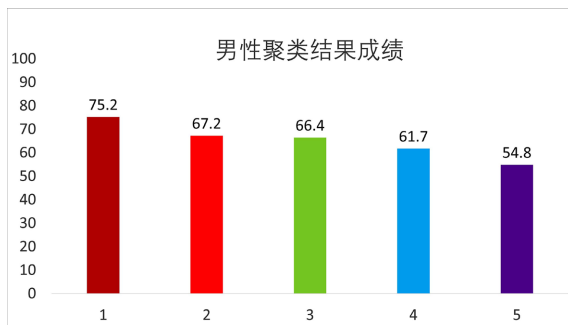


Figure 4. Male clustering results

图 4. 男性聚类结果

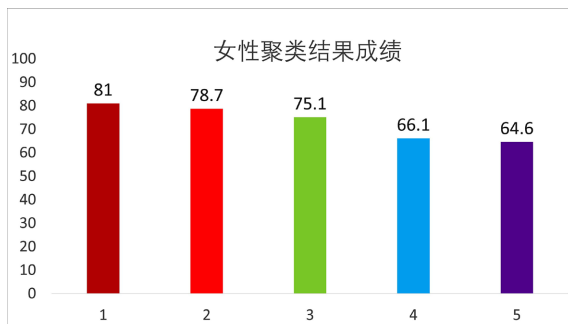


Figure 5. Female cluster results

图 5. 女性聚类结果

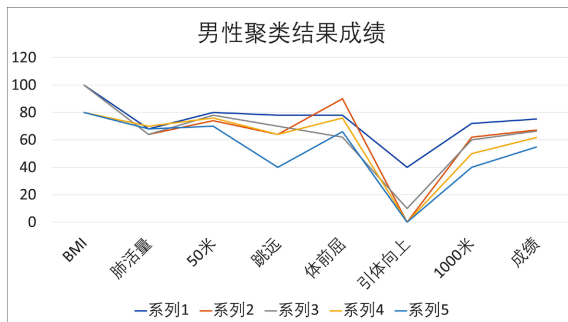


Figure 6. Line graph of the clustering results of male exercise levels

图 6. 男性各项运动水平聚类结果折线图

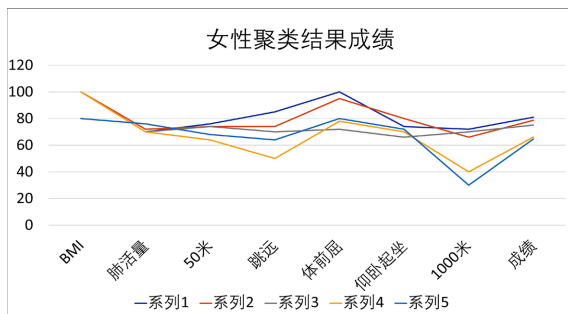


Figure 7. The line graph of the clustering results of female sports levels

图 7. 女性各项运动水平聚类结果折线图

通过 K-medoids 聚类分析, 我们依照现有通用的运动表现综合评价标准, 重新计算得出了更加符合上海理工大学学生的体质分类数据, 有助于教师根据上海理工大学学生具体身体素质以及不同的分类水平, 更好的制定适合学生的精细化运动模式, 根据不同水平学生的具体情况, 提供符合其健康需求的运动辅助, 帮助学生实现运动目标实现因材施教的差异化教学; 同时, 更加适宜的体质分类模式, 也有助于学生更加全面、客观地了解自我健康情况, 清楚自我体质水平定位, 帮助学生科学地明确、制定、实现属于自己的运动目标, 使运动处方更加适宜学生体质需求, 运动取得最大收益, 助力学生体质不断进步, 同时通过实时数据的更新, 可以掌握最新的学生体质情况, 方便教师及时修改教案以符合学生健康需求; 在疾病预防方面, 也有助于学生时刻监管自己的潜在健康风险, 使学生了解自己的体质水平以及相应的潜在健康风险, 提前制定科学的运动处方, 提高运动表现, 改善身体素质, 帮助学生尽早发现并开展各项运动健体从而尽早规避健康风险。

3.6. 基于 K-Medoids 算法的大学生体质分类评价模型的应用

3.6.1. 掌握学生健康情况, 为学生提供个性化运动处方

大学阶段, 学生往往面临着学业与工作的双重压力, 因而容易出现作息不规律, 饮食不健康, 缺乏体育锻炼等健康问题, 因此及时有效的体质监测手段尤为重要, 其帮助学生客观认识自己所面临的健康问题, 同时横向比较其他体质水平人群的运动表现, 更加具有针对性的进行体育锻炼, 了解自己的健康需求并制定切实合理的运动处方, 解决诸如像减脂、增肌、健美等多样化的健康需求, 使运动处方更加符合每个学生的需求及运动能力, 让学生有效、及时地取得最佳运动收获, 促进学生体育锻炼的积极性。对于教师, 该评价模型提供了一种动态的体质水平监测模式, 教师能够更加全面、具体、及时地了解学生体质水平分布及运动表现能力范围, 从而制定更加适合学生健康需求的教学计划, 从而保证教学的科学性、有效性, 为教师识别个体学生体质与群体差异, 了解学生群体体质水平提供科学、客观的度量工具, 促进教学内容的精细化、科学化发展。

3.6.2. 拓展体质分类标准, 结合实际情况进行不同群体的体质水平监测

目前, 各高校大多依照《国家学生体质健康标准》开展学生体质测评工作, 虽然《国家学生体质健康标准》具有高度科学性、客观性、普遍性, 但由于中国幅员辽阔, 不同大学所处环境往往有较大差异, 此外, 对于不同特色的大学, 其专业设置同样可能影响学生的体质健康发展, 针对总体学生体质水平较高的高校来说, 《国家学生体质健康标准》的标准可能低于学生普遍水平, 对于学生客观评价体质健康, 提升体质水平参考性较弱, 而对于总体学生体质水平较低的高校, 《国家学生体质健康标准》的标准可能高于学生平均运动能力, 从而总体呈现出较弱的体质水平。基于 K-medoids 算法的大学生体质分类评价模型的推广应用, 有助于不同高校在《国家学生体质健康标准》指导下, 通过运用模型实际研究高校具体学生体质, 开发出属于特定高校的个性化学生体质分类模型, 使学生体质分类更加贴近学生实际运动表现, 从而帮助教师与学生客观认识学生体质水平, 制定科学合理的运动目标与运动处方。此外, 该模型除了应用于高校学生体质水平分类以外, 还可应用于诸如不同体育班学生体质水平研究与比较, 通过对不同体育班学生进行聚类分析, 得出不同体质水平具有代表性的数据, 从而比较不同体育运动的运动能力训练特性; 将其运用于不同专业学生体质比较, 分析专业学习对于学生体质的影响等方面, 具有广泛的应用范围与实践价值。

3.6.3. 丰富健康信息渠道, 精准服务目标学生

通过对学生进行体质划分, 教师可以轻松针对不同水平的学生开展运动处方制定, 通过教学课程、健康平台搭建、健康个性化分析, 精准推送不同体质水平人群所需的健康信息, 运动处方, 满足不同人

群的健康需求,提供更加个性化的定制服务,学生也可从不同板块的健康信息中,找到提升体质水平的方法,激发学生健康运动的积极性。同时,平台的搭建帮助收集海量大数据,通过更大的群体模型分析,得出更加符合现实,具有科学依据的体质水平结论,通过大数据赋能学生体质健康促进工作以及体质健康干预过程。

4. 结语

本文建立了基于 K-medoids 算法的大学生体质分类评价模型,以《学生体质健康标准》为基础,通过收集上海理工大学 2022 年大学生体质健康测试数据,通过设计聚类算法模型,依据大学生各项运动表现水平将其体质划分为 5 类。通过研究国家发布的《学生体质健康标准》,与上海理工大学体育部合作,得到全校学生的体侧数据,剔除异常数据,按照男女进行分类,进行分析。高校教师在进行体育教学时可以根据不同学校的具体情况,建立属于本校学生的具体体质水平划分标准,并与《国家学生体质健康标准》相结合,运用大数据针对学生体质展开分析,使学生体质评估工作更加准确,教师更容易开展个性化教学设计工作。同时高校学生可以更加清晰认知自身存在的体质问题,激励学生改善个人体质水平,促进学生德智体美劳全面发展,加快体教融合,从而推动青少年文化学习与体育锻炼协调发展。

参考文献

- [1] 吕敬惠,袁愈新,王冬. 新型冠状病毒肺炎疫情下大学生心理健康状况及影响因素分析[J]. 南方医科大学学报, 2020, 40(2): 171-176.
- [2] 教育部网站. 教育部关于印发《国家学生体质健康标准(2014 年修订)》的通知[EB/OL]. http://www.moe.gov.cn/s78/A17/twys_left/moe_938/moe_792/s3273/201407/t20140708_171692.html, 2022-08-24.
- [3] 马德浩. 新时代我国高校体育发展的使命、挑战与对策[J]. 体育学刊, 2018, 25(5): 5-12. <https://doi.org/10.16237/j.cnki.cn44-1404/g8.2018.05.003>
- [4] 吕宏蓬,王天芳,张靖,等. 913 例大学生中医体质分类与卡特尔 16 种人格因子的相关性研究[J]. 中医杂志, 2022, 63(10): 962-967. <https://doi.org/10.13288/j.11-2166/r.2022.10.013>
- [5] 安义德,张霏,张欣,等. 中等强度运动处方对女大学生体质干预效果分析[J]. 中国学校卫生, 2022, 43(10): 1500-1504+1508. <https://doi.org/10.16835/j.cnki.1000-9817.2022.10.014>
- [6] 《中国人群精准运动处方研制与运动处方库建设》课题组,任弘,赵元慧. 主动健康理念下中国人群精准运动处方研制与运动处方库建设[J]. 北京体育大学学报, 2023, 46(11): 102-117. <https://doi.org/10.19582/j.cnki.11-3785/g8.2023.11.010>
- [7] 高峰,刘阳,吴印林. 新时代背景下高校体育教学模式创新研究——评《高校体育教学翻转课堂模式构建研究》[J]. 应用化工, 2023, 52(8): 2506. <https://doi.org/10.16581/j.cnki.issn1671-3206.2023.08.024>
- [8] 潘立成,黄海滢,张亮. 大数据背景下学生体质健康分类促进模式的应用研究[J]. 体育科技文献通报, 2023, 31(6): 165-167+261. <https://doi.org/10.19379/j.cnki.issn.1005-0256.2023.06.042>
- [9] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 67.
- [10] 李乐,王斐. 基于层次策略的半监督 K-medoids 算法研究[J]. 计算机应用研究, 2021, 38(5): 1387-1392. <https://doi.org/10.19734/j.issn.1001-3695.2020.05.0122>
- [11] (2020) Engineering; Studies from University of Azad Jammu and Kashmir Describe New Findings in Engineering (K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir). *Journal of Engineering*, 3751.
- [12] 邢聪. 学校体育课程“运动处方教学模式”的设计与构建[J]. 中国学校卫生, 2021, 42(9): 1288-1292. <https://doi.org/10.16835/j.cnki.1000-9817.2021.09.004>