

面向大数据人才培养的算法实践能力提升改革 ——以《机器学习基础实践》为例

胡 波¹, 邓 欣¹, 孙开伟¹, 杜雨露¹, 胡玲碧²

¹重庆邮电大学计算机科学与技术学院, 重庆

²重庆工程学院软件学院, 重庆

收稿日期: 2024年4月27日; 录用日期: 2024年5月24日; 发布日期: 2024年5月31日

摘要

机器学习基础实践课程在大数据人才培养过程中发挥着不可替代的重要作用, 因此探究该课程的教育教学改革具有重要的意义。文章针对课程教学中存在的问题, 提出了面向大数据人才培养的算法实践能力提升改革方案。方案通过优化教学内容, 完善课程资源, 丰富教学手段以及提升考评机制等举措不断强化学生的实践创新能力。通过本文的研究, 有望构建全方位多维度的机器学习实践课程体系, 收获更好的教学效果, 为大数据及其他新兴学科的实践教学提供有价值的参考, 对提高人才培养质量具有积极作用。

关键词

大数据, 机器学习, 实践课程, 教学改革

Reform for Improving Algorithmic Practice Capabilities in Big Data Talent Training —Taking “Basic Practice of Machine Learning” as an Example

Bo Hu¹, Xin Deng¹, Kaiwei Sun¹, Yulu Du¹, Lingbi Hu²

¹College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing

²School of Software, Chongqing Institute of Engineering, Chongqing

Received: Apr. 27th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

Abstract

The foundational practical course of machine learning plays an irreplaceable and significant role

文章引用: 胡波, 邓欣, 孙开伟, 杜雨露, 胡玲碧. 面向大数据人才培养的算法实践能力提升改革[J]. 教育进展, 2024, 14(5): 1380-1387. DOI: 10.12677/ae.2024.145852

in the cultivation of talents in big data. Therefore, investigating the educational and teaching reforms of this course holds important significance. This article puts forward a reform scheme for enhancing algorithmic practical abilities aimed at cultivating talents in big data. The scheme continuously strengthens students' practical innovation capabilities by optimizing the teaching content, improving course resources, enriching teaching methods, and enhancing the evaluation mechanism, addressing the problems existing in the course's teaching. Through this research, it is anticipated to construct a comprehensive and multi-dimensional system for the practice of machine learning courses, achieving better teaching outcomes. This will provide valuable references for the practical teaching of big data and other emerging disciplines, and play a positive role in improving the quality of talent cultivation.

Keywords

Big Data, Machine Learning, Practical Courses, Teaching Reform

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

2015年9月5日，国务院颁布《促进大数据发展行动纲要》，开始系统谋划大数据发展工作。为贯彻执行该行动纲要以及国家“十三五”规划中的大数据发展战略，2016年12月18日，工业和信息化部发布了《大数据产业发展规划(2016~2020年)》，进一步明确数据是国家基础性战略资源，全面推进大数据发展，加快建设数据强国。在大数据人才培养方面，国家也提出了总体要求，即“建立适应大数据发展需求的人才培养和评价机制。加强大数据人才培养，整合高校、企业、社会资源，推动建立创新人才培养模式，建立健全多层次、多类型的大数据人才培养体系”。然而，据人力资源和社会保障部2020年发布的《人工智能工程技术人员就业景气现状分析报告》，我国人工智能人才缺口超过500万，国内的供求比例为1:10，供需比例严重失衡。同时，工业和信息化部人才交流中心发布的《人工智能产业发展报告(2019~2020年版)》指出机器学习在人工智能产业整体需求岗位中的占比达39.1%，但相关技术方向的人才极度稀缺，人才供需比仅为0.23，有效供给严重不足。而高等院校作为承担为国家培养大数据人才重任的第一“桥头堡”，应该尽快建设起相应的大数据人才培养方案、课程体系等。

2016年，教育部发布的《2015年度普通高等学校本科专业备案和审批结果》中，首次增设数据科学与大数据技术专业，北京大学、对外经济贸易大学、中南大学成为第一批成功申请到“数据科学与大数据专业”的学校。重庆邮电大学于2018年获批数据科学与大数据技术专业并开始招生，是国内第三批获批高校之一(当年共248所学校获批)。数据科学与大数据技术专业作为新开设的本科专业，在不同高校根据各自教学资源和办学条件制订相应的人才培养目标。虽然各具特色，总体来看，该专业主要旨在培养具备信息科学、数理统计和数据科学基础知识与基本技能的学生。他们将掌握大数据采集、预处理、存储、处理、分析、可视化和数据安全等技术，具备一定的数据科学研究能力以及数据科学家岗位所需的基本素养。他们将能够胜任数据分析与挖掘算法研究以及大数据系统开发的研究型和技术型人才[1]。

《机器学习基础实践》是数据科学与大数据技术专业的一门专业实践课程。课程内容涵盖：线性模型、决策树模型、神经网络模型、支持向量机模型、集成学习方法、贝叶斯分类模型、半监督学习等[2]。课程通过大量的实践案例，让学生在实践中理解机器学习的基本理论知识，掌握应用机器学习算法解决

实际问题的步骤和方法，学会选用合适的机器学习方法进行问题解决方案的设计和实现，为后续专业课程的学习打下坚实的理论和实践基础。毫不夸张的说，机器学习实践课程在大数据分析、自然语言处理、计算机视觉等人工智能应用中发挥着关键的基础性作用，见图 1。

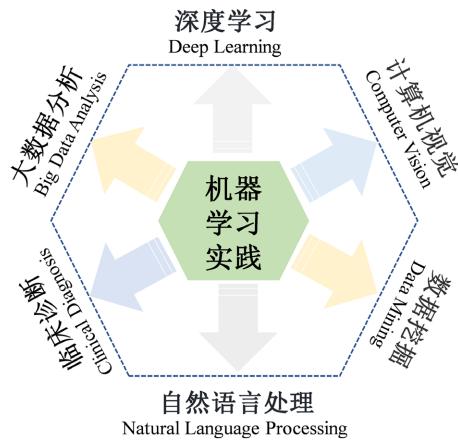


Figure 1. The important position of the machine learning practical course
图 1. 机器学习实践课程的重要地位

因此，探索面向大数据人才培养的算法实践能力提升改革——以《机器学习基础实践》为例十分必要且具有重要意义。通过本论文的研究，有望构建全方位多维度的课程体系，收获更好的教学效果，为大数据及其他新兴学科的实践教学提供了有价值的参考，对提高人才培养质量具有积极作用。

2. 现有问题分析

数据科学与大数据技术专业面向产业培养具备大数据系统平台研发、大数据应用系统开发和大数据分析能力的创新型数据人才。大数据技术涉及计算机科学、统计学和数学等相关领域的深度融合，是一个典型的跨学科和领域的交叉学科。行业对大数据人才的要求不仅需要具备丰富的理论知识，还需要具备更强的实践能力。因此，实践教学是大数据人才培养体系中的重中之重。《机器学习基础实践》是数据科学与大数据技术专业的一门专业实践课程，为后续专业课程的学习打下坚实的理论和实践基础。然而，本门课程在教学过程中面临着以下几个方面的挑战与不足：

2.1. 教学内容陈旧，缺乏前沿技术

传统“机器学习基础实践”课程教学根据大纲内容要求，主要以数学模型概念讲解、公式推导和实践习题为主，缺乏对机器学习技术的发展过程、历史背景等方面的详细介绍，教学内容局限于机器学习模型的理解、推导和应用，缺乏课外知识的拓展，教学内容单调且枯燥，课堂氛围活跃度不高，学生无法感受机器学习技术发展对实际生活的影响和学习机器学习技术的重要性。同时，在上机实践过程中，传统“机器学习”课程内容主要关注回归模型、决策树模型、支持向量机模型、贝叶斯分类器、集成学习模型等(见图 2)，要求学生实现模型的编程和预测过程[3]。经典的机器学习案例实践能够在一定程度上帮助学生理解模型概念和实现方式，但与学生的专业背景和现实生活却存在一定距离，学生无法切身体会掌握机器学习知识对其生活和工作产生的具体作用。因此，脱离实际生活的教学内容最终将使学生的学习兴趣随着教学进度的推进而消磨殆尽。

近年来，以卷积神经网络、长短期记忆网络、生成对抗网络等深度学习模型为主的机器学习技术已经广泛应用于人脸识别、异常检测、自然语言处理等方面，相比传统的机器学习模型，其在精度和稳定

性方面更具有优势。然而，在课程教学实践中，考试内容常选择线性回归模型、决策树、反向传播神经网络、支持向量机等经典模型作为主要考点，学生为获取更好的考试成绩而花费大量时间在传统模型基础概念复习和公式推导，无暇顾及机器学习前沿技术的学习，导致学生对前沿技术不敏感，抑制了学生的自主学习热情，不利于培养学生的终身学习能力。

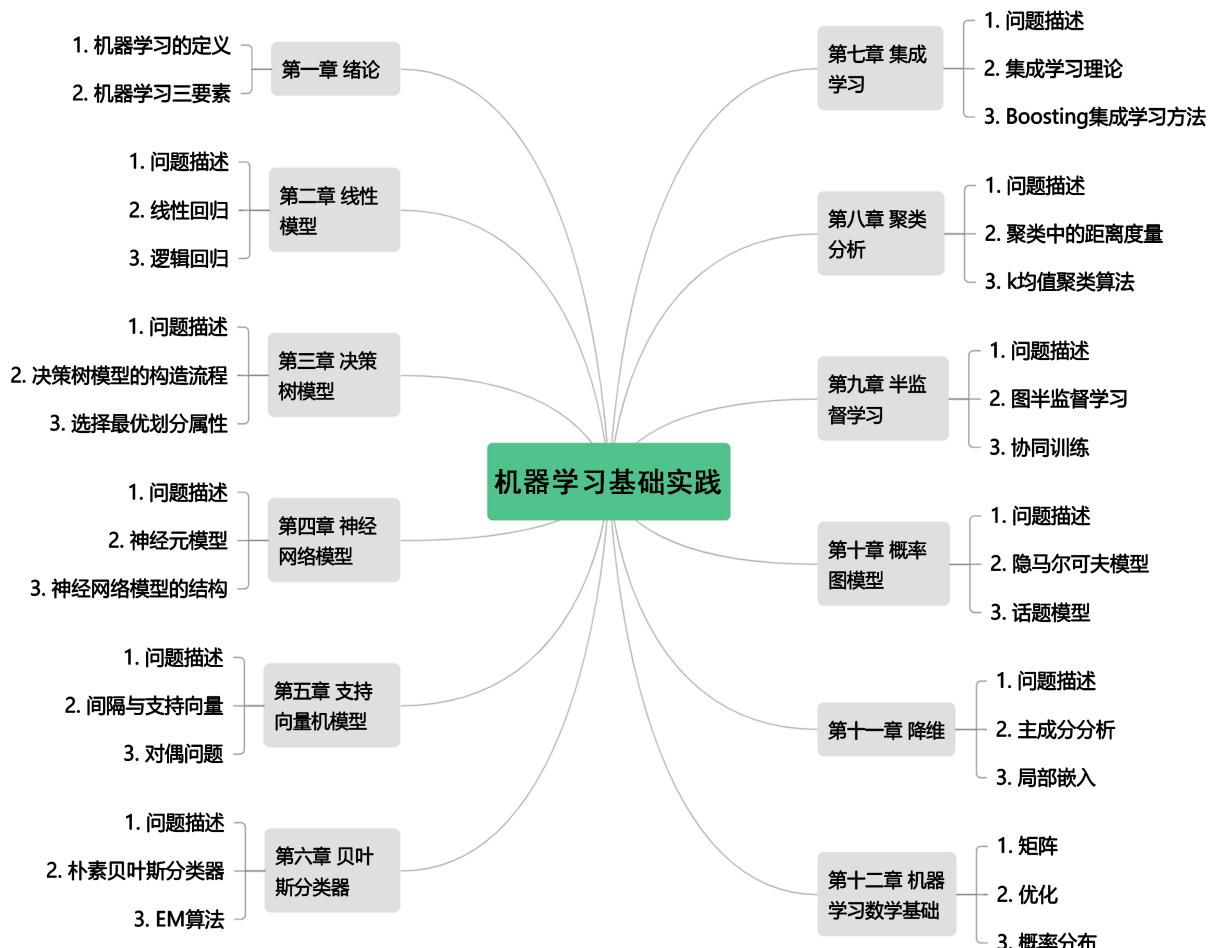


Figure 2. Chapter content of “Fundamentals of Machine Learning Practice”

图 2. 《机器学习基础实践》章节内容

2.2. 课程资源单薄，体系不够完善

对于课程资源的理解，总体来说，就是一切能够运用到教学活动中的各种条件和材料。能促进教学活动更好的开展，是提高教育教学水平和科研水平，是保证人才培养质量的关键，如图 3 所示。然而，现有课程资源单薄，体系不够完善，主要体现如下：

首先，《机器学习基础实践》课程教学 PPT 制作较为松散，章节之间的逻辑连贯性不强，且 PPT 内容粗糙，有很大提升空间；其次，实验指导书案例较为陈旧，分析内容较少，学生难以进一步理解所学内容，容易导致刻板重复。此外，实验指导书中代码的呈现对初学者并不友好，学生往往直接复制粘贴到软件中进行编译运行，大量可能的语法错误打击学生的自主学习积极性；再次，网络学习资源非常有限，仅有吴恩达机器学习系列课程、周志华机器学习系列课程[4]等相关视频资源，然而此类课程主要面向机器学习理论课教学，与《机器学习基础实践》的教学目标和定位还有较大差异。因此，学生花费时

间后收获往往有限。总之，当前严重缺乏适用于大数据专业的视频教学课程，这严重阻碍了学生的进一步学习和进步；最后，上述课程资源之间的关联性不强，各自发挥着非常有限的作用，没有形成课程教学资源体系，导致学生收益较小。

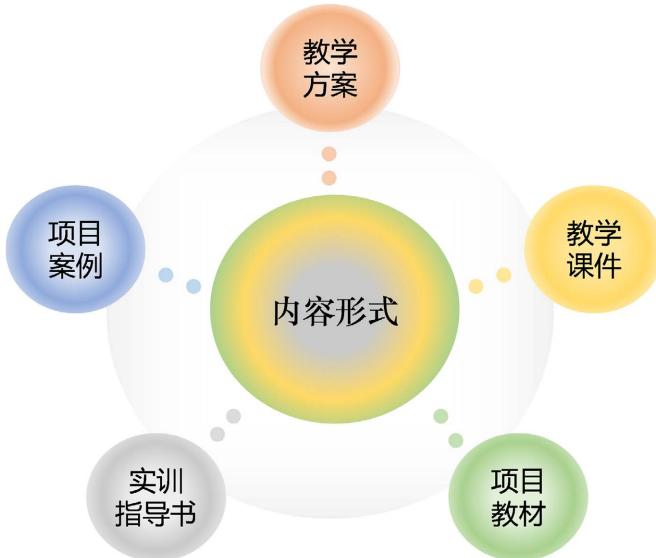


Figure 3. Overview of course resources
图 3. 课程资源概况

2.3. 教学形式单一，方式缺乏创新

目前，面向本科生的课堂教学主要还是借助多媒体和 PPT 展示教学内容，以教师讲授、学生听讲为主要形式的教学模式。首先，由前文分析可知机器学习基础实践的理论知识多，采用当前的常规教学方式，难以激发处于被动听课状态的学生积极思考，造成他们缺少积极主动的思维意识，缺少积极主动的思维心态；其次，机器学习基础实践内容复杂、算法难以理解，仅仅依靠多媒体和板书方式做不到形象的分析，不容易帮助学生掌握，他们会逐渐失去听课的动力，无法保持听课的专注力，沉闷单调的课堂气氛必然影响教学效率和效果；再次，教师在课堂上大多使用电子课件进行讲授，容量大、速度快，使得学生对所学内容记忆不深刻，对该门课程的认识不到位，对于课程能否有助于学生的后续发展有所疑虑，从而导致学生的抵触情绪；最后，给学生留下“机器学习枯燥无趣”的印象，对该课程失去信心和兴趣。面对以知识灌输为特征的传统教学方式，教学形式单一，方式缺乏创新，收效甚微[5]。

2.4. 课程考核落后，过程体现不足

考评是任何课程教学活动中不可或缺的重要环节，它具有双重功能，既可以检验成效，即评价教师的教学效果，检验学生的学习结果，又可以促进改革，改进教师的教学内容和教学方法，调整学生的学习侧重点和学习方法[6]。当前本课程的考核方式包括平时考勤、课堂考核、实验报告。对应的评分标准为：1) 平时考勤：考勤不参与最终评分，缺勤次数超过 40% (缺勤 4 次及以上)，该课程最终成绩为零；2) 课堂考核：课堂考核占总分的 70%，根据每次课程的内容的完成情况，由任课教师在课堂进行评分；3) 实验报告：实验报告占总分 30%，按照评分细则进行，评分细则见表 1。

然而，现有课程考核落后，缺乏过程体现，主要分析如下：1) 虽然现有考核中考虑了平时表现，但“考勤不参与最终评分”，也就是对学生行为不构成约束，仅通过“缺勤次数超过 40% (缺勤 4 次及以上)，

该课程最终成绩为零”来要求学生，这就容易导致学生只重到教室，不重学知识，使学生觉得只要到教室签到，就可以满足平时考核，而平时的实验认真与否，质量高低可以不用那么重视；2) 课堂考核仅仅是对课堂课程实验情况做直观评价，而没有联系学生自身情况，比如是否较之前有所进步，是否有帮助其他同学，协作沟通能力等基本素养。

Table 1. Grading criteria for experimental reports
表 1. 实验报告评分标准

考核内容	90~100 分	75~89 分	60~75 分	小于 60 分
实验作业完成进度 (权重 0.1)	提前完成	按时完成	延时完成	补交
实验报告文档 (权重 0.7)	文档报告充实，实验分析、设计、测试正确，文档不少于 5 页	文档报告较充实，实验分析、设计、测试基本正确，文档不少于 5 页	文档报告不够充实，实验分析、设计、测试错误较多，文档不少于 5 页	文档报告不充实，内容错误较多，文档篇幅不够
实验报告总结与分析(权重 0.2)	能总结实验结果给出总结，分析实验结果的原因，见解独到	能总结实验结果给出总结，有一定见解	能总结实验结果给出总结	无实验结果总结

3. 课程教学改革措施

3.1. 制定合理教学内容，新旧技术齐头并进

针对现有课程建设“教学内容陈旧，缺乏前沿技术”这一问题，本文开展重构教学内容的研究，使之科学合理，且注重新旧知识的有机结合。第一，《机器学习基础实践》与大数据专业课程群存在知识点交叉，在有限的课时内所选择的教学内容应该合理适量，避免重复；第二，经典知识点应该在调研学生基础和特点的基础上，有重点地选择代表性算法呈现机器学习的知识体系；第三，前沿知识点的讲授重在激发学生的学习兴趣，激发学生的求知欲，所以难度应当适中，如近年来发展火热的深度学习技术(比如，深度伪造、ChatGPT、AIGC 等)，需要适当介绍神经网络知识作为先导。

3.2. 加强课程资源建设，不断完善体系结构

针对现有课程建设“课程资源单薄，体系不够完善”这一问题，本文开展课程资源建设计划，完善课程资源体系。本课的课程资源建设分为线下资源和线上资源。线下资源主要由老师做好理论课讲授的内容，包括 PPT、视频、教案、实验指导书等。线上资源的建设，首先要求老师寻找合适的慕课平台，线上线下相互结合，把慕课当作线下的补充，利用慕课平台为学生提供课前预习的材料或课后拓展巩固的资料等，同时还可以基于慕课进行翻转课堂教学，要求教师精心设计教学活动的各个环节，如组织课堂教学流程，组织学习讨论，课堂学习评比，以调动学生学习的自觉性。其次，除了现有的网络免费公开课(慕课、网易公开课、国内外著名高校公开课)，还将针对我校学生的特点建设自己的线上学习平台。

3.3. 丰富教学手段形式，围绕学生开展创新

针对现有课程建设“教学形式单一，方式缺乏创新”这一问题，本文开展教学形式和手段的创新与丰富研究。不同于传统的被动学习方式，教师应该对教学倡导主动学习。第一，教师应当要求学生强化课前预习，课堂上把握好教学节奏，适当时间内讲解完重点难点后，设置恰当的提问解答互动环节，让学生加强思考增强乐趣，同时在此过程中也活跃了课堂气氛。第二，课后教师可以跟学生进行互动交流，这样不仅可以了解学生对课程的想法和建议，作为教学改进的依据，还可以增进师生感情，提升学生对

教师的认可度,从而提升学生学习兴趣[7]。第三,教师可以鼓励学生利用丰富的网络资源自行查找相关文献学习相关知识点,参考公开数据和程序,推荐国际会议论文集等,指导学生参与各种线上线下的研讨交流活动,比如参与 Valse 论坛,浏览小木虫讨论,中国知网等,提升学生学习的主动性和积极性。第四,鼓励学生组队参加大数据竞赛,并分享解决方案,以此提升讲授学生自信,增强其余学生的兴趣。

3.4. 丰富教学手段形式,围绕学生开展创新

针对现有课程建设“课程考核落后,过程体现不足”这一问题,本文开展全过程多维度的考核方式研究,力求从平时过程中把控学生的学习意识。第一,从课堂教学,课上实践,期末报告等多角度思考,掌握学生学习情况。第二,组织小组参与难度适中的竞赛项目,并以答辩的形式给与考核。项目实践可由教师根据课程内容和学生科研方向精心引入应用案例,设计层次化区别化的课题,让学生在项目中综合理解机器学习理论与实践[8]。第三,开展以专题汇报的形式,全面考察学生学习情况。最终,学生的成绩包括多个教学环节的综合评定,如表 2 和图 4 所示。

Table 2. Performance assessment items and proportion

表 2. 成绩考核项目及比例

类别	方式	权重(%)	总比例(%)
课堂教学	参与讨论	10	20
	专题分享	10	
课后实践	实验报告	10	50
	算法实现	20	
	项目答辩	10	
课程报告	大作业	30	30

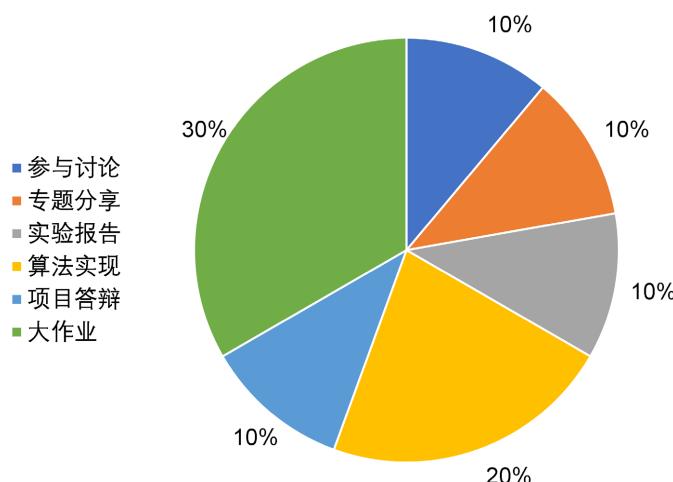


Figure 4. Performance assessment items and proportion

图 4. 成绩考核内容及比例

4. 结语

数据不仅是信息时代的核心资产,更是支撑科技创新、经济发展和社会进步的重要基石。当前,大数据对全球各国的行业发展、社会生活、以及政策制定都产生了重大影响。机器学习作为主要用于数据

分析的工具,对大数据及其相关专业人才的培养具有十分重要的意义。本文围绕大数据人才培养的算法实践能力提升改革,针对当前机器学习基础实践课程教学中的突出问题,提出了多维度全方位的解决方案,涵盖了教学内容、教学手段和教学评价等课程核心内容。通过本文的研究,有望对国内高校大数据及其相关专业的实践课程教育教学改革提供借鉴参考。

基金项目

本文得到重庆邮电大学教育教学改革研究项目(XJG23105, XJG21224, XJG23229)、重庆市高等教育教学改革项目(213156, 233210)的资助。

参考文献

- [1] 孙开伟, 邓欣, 王进. 新工科背景下数据科学与大数据技术专业实践教学体系研究[J]. 高教学刊, 2023, 9(14): 5-8.
- [2] 朱灵龙, 曹海啸, 阚希, 等. 新工科背景下应用型本科高校数据科学与大数据技术专业建设研究[J]. 信息系统工程, 2024(3): 166-169.
- [3] 楚晓琳, 王江盼, 丁雪. 基于 OBE 理念的“机器学习基础”课程教学改革探索[J]. 科教导刊, 2024(3): 88-90.
- [4] 张俊丽, 姚香秀, 程茜, 吴睿. 基于 OBE 理念的数据科学与大数据技术人才培养创新研究与实践[J]. 大学教育, 2022(4): 185-187.
- [5] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [6] 余波. 基于成果导向的“机器学习”课程教学改革[J]. 工业和信息化教育, 2022(8): 24-28.
- [7] 冯岩, 尤磊, 李健, 等. 新工科背景下人工智能课程的教学改革[J]. 福建电脑, 2022, 38(4): 118-120.
<https://doi.org/10.16707/j.cnki.fjpc.2022.04.031>
- [8] 蒋磊, 张丽, 闫俊, 等. 机器学习实践课程教学改革探讨[J]. 教育现代化, 2020, 7(56): 76-79.
<https://doi.org/10.16541/j.cnki.2095-8420.2020.56.022>