

面向创新能力培养的机器学习与数据挖掘 教学改革探索^{*}

王盛邦^{1,2} 林惊^{2**}

1 广州新华学院信息与智能工程学院, 广州 510520;
2 中山大学计算机学院, 广州 510006

摘要 针对机器学习与数据挖掘课程教学中创新能力培养不足的问题, 提出以创新为导向的教学改革思路。首先重构课程结构, 将理论教学、实践训练与创新引导相结合; 然后提出以问题为导向、项目为载体的“五段式”引导教学方式, 以开放、多元的实验与开发平台为支持, 制定能够反映学生创新能力的多维评价机制。实践表明, 该改革思路能有效提升学生的创新思维水平和解决实际问题的能力。

关键字 机器学习, 数据挖掘, 教学改革, 创新能力培养, 项目驱动, 实践教学, 评价体系

Exploration of Teaching Reform in Machine Learning and Data Mining for Cultivating Innovation Capabilities^{*}

Wang Shengbang^{1,2} Lin Jiang^{2**}

1.School of Information and Intelligent Engineering, Guangzhou Xinhua University,
2.School of Computer Science and Engineering Sun Yat-sen University Guangzhou, China;

Abstract—Faced with insufficient cultivation of innovation ability in the teaching of Machine Learning and Data Mining, this study proposes an innovation-oriented reform blueprint. The curriculum is first restructured to fuse theoretical instruction, practical training, and innovation mentoring. A five-stage, problem-based, project-driven pedagogical model is then introduced, underpinned by an open and diversified experimental development platform, together with a multi-dimensional assessment system that captures students' innovative capacity. Empirical results demonstrate that the reform significantly enhances students' creative thinking and their ability to tackle real-world problems.

Keywords—Machine Learning; Data Mining; Teaching Reform; Innovation Cultivation; Project-Driven; Practical Teaching; Evaluation System

1 引言

在 AI 的广泛领域中, 机器学习(ML)与数据挖掘技术(DM)的突破与应用至关重要。ML 是其核心分支, 而 DM 则是从数据中提取价值的技术^[1]。ML 与 DM 是高校计算机专业尤其是 AI 专业重要的专业课程。然而, 该课程在教学中存在创新能力培养不足的突出问题。一方面, 教学内容偏重算法理论, 与实际应用场景之间缺乏衔接, 导致学生虽掌握基础知识, 却不懂如何进行创新^[2]。另一方面, 实践教学环节多停留于纸上实验, 难觅面向真实问题的开放性实践项目, 不利于激发学生探索未知与解决复杂工程问题的能力^[3]。同时, 课程评价体系仍以理论性笔试为主, 无法衡量学生在创新思维、系统设计与综合实践等方面的真实水平^[4]。此外, 课程内容更新缓慢, 教材明显滞后于技术发展。

再加上学生对行业前沿和实际应用了解甚少, 缺失创新视野, 也很难把不同领域的知识融会贯通。

为提高教学成效, 一些同行进行了改革尝试。文献^[5]提出层次化认识模型, 说明该方法可提升机器学习课程教学效果。在教学方法上, 文献^[6]依托翻转课堂突出学生实践主体性, 着力培养其应用机器学习算法解决实际问题的综合能力。文献^[7]的创新点在于将

“头脑风暴”模式融入培养过程, 并配套建立完善的考核评价体系, 以推动数据分析与挖掘课程的深度改革。文献^[8]以决策树分类模型为例, 阐述如何进行课程内容重构、课程思政融合和分层资源构建。以上这些探索均取得一些成效。与之不同的是, 本文以创新能力培养为核心目标, 对 ML/DM 课程进行系统性教学改革, 既能及时应对技术发展需求(现实意义), 又能深化相关教育理论研究(理论价值)。

2 ML/DM 教学中创新能力培养的瓶颈分析

*基金资助: 本文得到国家级实验教学示范中心联席会计算机学科组规划教材建设课题(CPEC2025-P-05)资助。

**通信作者: 林惊, 男, 教授, linliang@ieee.org。

2.1 教学理念滞后

课堂仍以教师讲授为主，学生被视为知识的被动接受者，而非主动的探索者和创造者^[9]。一些教师对“创新”内涵理解不足，教学中缺乏系统性的创新能力培养机制设计。所谓的案例教学实际上只是对成熟解决方案的展示，缺乏引导学生主动发现问题、提出假设、设计方案并验证的探究过程。讨论、辩论、头脑风暴等激发创造性思维的方法应用不足，甚至压根没有用到。

另一方面，与学科前沿脱节，这是较普遍情况。过于依赖教材使得课程内容明显滞后于技术发展，对AutoML、可解释AI(XAI)、联邦学习、图神经网络(GNN)等前沿方向及其蕴含的创新点涉及较少。课程内容片面注重算法模型(如SVM、决策树、神经网络、聚类算法等)，而缺乏对算法设计思想、适用边界、组合创新以及在不同领域应用创新的深入探讨。

2.2 实践深度与开放性不足

实践普遍浅层化，实验项目多为使用成熟库(如Scikit-learn、TensorFlow/PyTorch)调用API处理清洗好的标准数据集(如Mnist、Iris)，任务目标明确、步骤固定。学生无需深入理解数据特性、进行特征工程的创造性设计、应对数据噪声与缺失的挑战，也较少涉及模型选择、调参优化背后的系统思考和权衡。计算资源有限，缺乏接触真实、复杂、多源、有噪声的大规模数据集的机会。与企业实际需求、行业真实问题场景的连接薄弱，甚至没有。

2.3 评价机制因循守旧

现行评价体系过度强调对确定知识点和固定操作步骤的掌握(如期末笔试占比过高，普遍达到60%)，对学生在项目过程中展现的对问题定义的新颖性、解决方案的独创性、克服技术难点的毅力、实验结果分析的批判性思维以及团队协作中的创造性贡献程度缺乏有效的识别和激励^[10-11]。

3 面向创新能力培养的教学改革核心方案

3.1 课程体系重构

针对上述瓶颈，我们构建了以“激发创新意识、夯实创新基础、锤炼创新实践、塑造创新人格”为核心理念的教学改革体系。图1是所构建的“理论-实践-创新”金字塔模型。

理论层立足扎实理论，理解本质。其核心内容包括概率统计、线性代数、优化理论基础；经典监督/无监督学习算法(LR、SVM、DT/RF、K-Means、PCA)

的数学本质、假设前提、推导过程与适用边界；模型评估与选择准则等。

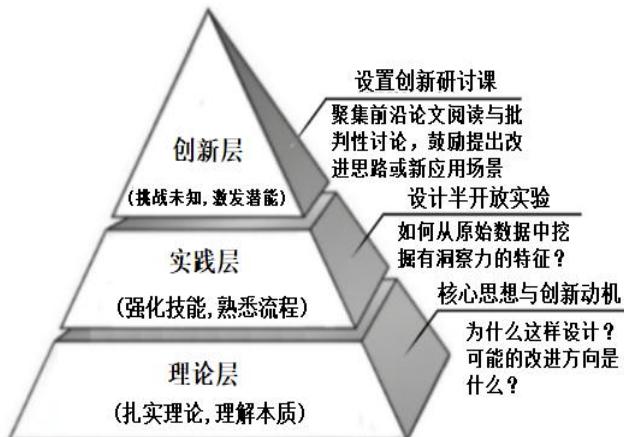


图1 “理论-实践-创新”的金字塔模型

理论层创新融入点：重点讲授算法设计背后的核心思想与创新动机(如SVM的最大间隔思想、集成学习的多样性价值)。引导学生思考“为什么这样设计？”、“可能的改进方向是什么？”。引入算法发展上的关键突破案例(如从感知机到神经网络)，剖析其创新思维。

实践层强化技能，熟悉流程。其核心内容包括数据预处理技术(清洗、集成、变换、规约)；注重特征工程的科学性(特征构建、选择、转换)；模型训练与调参(网格搜索、随机搜索、贝叶斯优化)、验证与部署基础；主流工具链(Python、Pandas、Scikit-learn、PyTorch/TensorFlow)。

实践层创新融入点：设计ML/DM半开放实验项目。例如，给定一个相对真实的数据集(如某电商用户行为日志)和一个模糊目标(如“提升用户体验”)，要求学生自行定义具体预测或分析任务(如用户流失预测、商品推荐、用户分群)，并自主完成从数据理解、任务定义、特征工程、模型选择与优化到结果分析的全流程。强调特征工程的创造性发掘(如何从原始数据中挖掘有洞察力的特征？)。

创新层勇于挑战未知，激发潜能。其核心内容包括前沿技术专题研讨(AutoML、XAI、迁移学习/联邦学习、时序分析、图学习等)；跨学科应用场景(如生物信息学、计算社会科学、智慧农业)；项目创新实践与学科竞赛等。

创新层创新融入点：通过“创新研讨课”和“综合项目实践”提升创新潜能。创新研讨课倡导学科前沿论文阅读与讨论，鼓励学生提出改进思路或新应用场景。综合项目实践要求学生自主选题(如科研课题、企业痛点或社会需求)，组队完成一个从问题定义、方案设计、系统实现到效果评估的完整创新项目。强

调方案的新颖性、可行性及潜在价值。

在教学改革的同时，注重将“立德树人”融入能力培养全过程。在理论教学中强调追求真理，严谨求实的科学精神；在实践教学中塑造精益求精的工匠品质；在综合创新中引导学生树立科技报国的使命感与自主创新的责任感，实现知识传授、能力提升与人格养成的深度融合。

3.2 教学模式革新

推行“问题驱动、项目贯穿、探究为本”的PBL（基于问题的学习）+模式[12]。

以实际（或模拟）问题为起点。在讲解算法时，通过实际生活的应用场景（或由ML/DM精心设计的模拟场景）展开。例如，在讲解分类算法前，先提出“如何根据用户历史行为精准预测其是否会点击某个广告？”、“如何从海量医疗影像中自动识别早期病变？”等实际问题。

以项目为抓手。将核心知识点应用于具有递进性挑战的项目中。例如在基础实践使用给定数据集和指定算法完成一个分类任务，重点练习数据预处理和模型评估；在进阶实践中选择最适合某类问题的算法模型，并进行深入的调优和特征工程尝试；在综合创新中自选数据集和难题，运用所学知识，提出创新性解决方案，撰写报告并参与答辩。

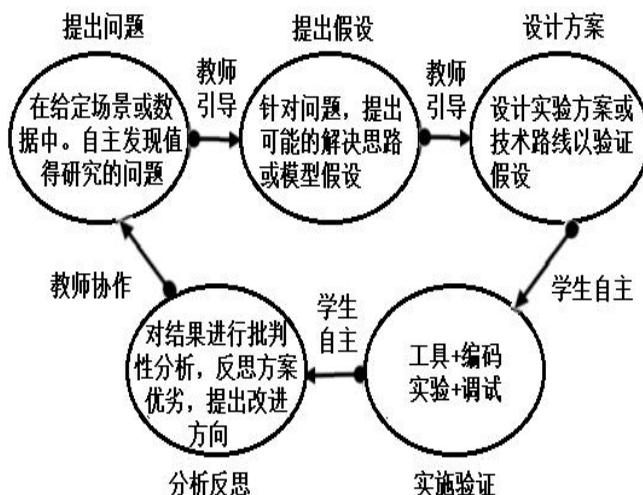


图2 “五段式”探究过程

强调探究过程。教师角色从“讲授者”转变为“引导者”和“协作者”。采用启发式、讨论式、案例式教学。

对给定场景（数据），实行“五段式”引导（即提出问题、提出假设、设计方案、实施验证、分析反思），其过程如图2。

3.3 开放协同的实践平台

强化校内平台。为满足ML/DM的实践场景，需要高性能计算集群和云实验平台的支持，并提供开放的实验环境、数据集访问权限、必要的软件工具和文献数据库等基础环境。

深化产学协同。引入企业专家参与课程共建，开展行业科普、前沿技术讲座、案例分析、项目指导等工作。建立校外实践基地，将企业需求转化为学生项目实践，让学生能够深入参与实际AI项目，增强对产业问题的理解与解决能力。

竞赛促学。积极组织并指导学生参与国内外高水平数据科学竞赛（数据分析的竞赛平台Kaggle、数据挖掘和知识发现年度竞赛KDD Cup、中国高校计算机大赛大数据挑战赛等），以赛促学、以赛促创。

利用开源与社区。鼓励学生利用GitHub等平台管理代码、学习优秀开源项目、参与开源社区讨论，提升项目效率和实践过程的可追溯性。

4 评价体系重构

摒弃“一考定乾坤”的传统模式，构建能全面反映学生创新能力成长过程的“多维度、过程性、重创新”的发展性评价机制。

4.1 多元化的评价机制

评价维度多元化主要包括知识理解与应用、实践技能、创新能力和协作与沟通四方面。

知识理解与应用指考核基础理论、算法原理掌握程度（可通过测验、作业）；实践技能指考核编程能力、工具使用熟练度、数据处理与建模能力（通过实验报告、代码审查）；创新能力指发现有价值问题的能力（项目选题报告）、解决方案的独创性与技术路线的合理性（项目方案设计书）、解决关键技术难题的能力（代码、文档），以及对结果的分析深度（项目报告、答辩过程）；协作与沟通则通过组员互评、教师评价及答辩表现，综合评估学生在团队项目中的协作与沟通能力，包括贡献度、沟通协调能力、报告撰写与展示能力。

4.2 多元化的评价方式

在评价方式上，可适当降低期末笔试权重（如降至总评的30%~40%），采用多元化的考评方式。以教师评价、学生自评、组间互评等相结合的方式，确保能全面衡量学生的学习成果。课程最终成绩的构成可设置为：理论考核(30%)、项目完成度(40%)、创新成果(20%)、平时成绩(10%)。

5 改革实践案例与初步成效

5.1 实践案例

以“医疗健康数据分析”创新项目为例。该案例以公开的糖尿病数据集（若条件许可可与附属医院合作，由其提供脱敏的糖尿病患者实际电子病历数据）。先由行业专家（如附属医院专家）进行糖尿病知识科普及最新研究成果介绍，然后学生自由组队（每队3人左右）开展选题调研和方案论证（含文献综述与技术选型），最终确立项目研究方向。表1列出了学生确定的3个选题。

表1 “医疗健康数据分析”创新项目

项目编号	项目名称	创新点
mldm01	基于多模态数据（结构化病历+文本描述）预测糖尿病并发症风险	融合文本特征（使用NLP技术）和传统特征，探索图神经网络（GNN）建模患者-并发症关系
mldm02	探究不同降糖药物组合对血糖控制效果的差异性	应用因果推断方法（如Propensity Score Matching）尝试从观察性数据中估计因果效应，而非简单关联分析
mldm03	开发一个轻量化的糖尿病风险自评工具（原型）	特征选择和模型轻量化设计，便于部署到移动端

项目确定后，需要对数据集进行数据处理（面临大量缺失值与噪声）、模型开发与迭代、结果分析与可视化，除了撰写报告，还须参与项目完成答辩。

5.2 初步成效

通过本轮教学改革实践，学生能力在多个方面呈现显著进步。

学生创新能力显著提升。项目选题多样性与前沿性明显增强，解决方案中尝试应用前沿技术（如GNN、

因果推断、AutoML工具）的比例有较大幅的提高。表明学生敢于挑战复杂问题和尝试新方法。

实践成果质量大幅提高。项目报告深度、代码规范性和项目完整性超过传统实验。

学习主动性与深度增强。PBL模式显著增强了学生的内在学习动机，成为解决项目中的挑战性问题的驱动力，具体体现在能主动查阅文献、学习课外知识（如新的算法、库）、在论坛提问、组内深入讨论等方面。对知识的理解由“事实性认知”（是什么）向“原理性理解”（为什么）深化，并拓展至“迁移性应用与创新”（还能怎么用）的深度。

团队协作与沟通能力得到较好锻炼。通过团队项目，学生在分工协作、技术讨论、冲突解决、成果整合与展示方面得到充分锻炼，同时也了解了行业需求和技术趋势。

表2是《机器学习与数据挖掘》课程改革前后三年的代表性数据，每班学生人数约60名。学生课程满意度调查中，“激发创新思维”、“提升解决问题能力”、“加强实践动手能力”等方面的评分平均提升超过了25%。教学满意度、创新项目参与率也有较大幅度增长，选择ML/DM相关方向作为毕业设计/研究的学生比例显著增加。图3更直观的体现这种变化，说明改革初见成效。

表2 教学改革创新能力指标

衡量指标	改革前	改革后
激发创新思维	10%	35%
提升解决问题能力	30%	56%
加强实践动手能力	35%	65%
教学满意度	72%	93%
创新项目参与率	35%	75%

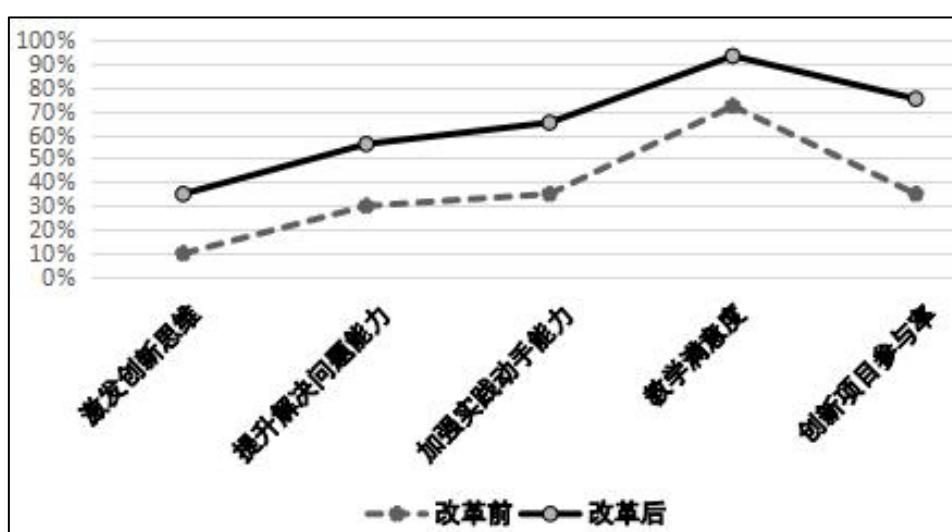


图3 教学改革创新能力指标变化

6 结束语

培养具备创新能力的机器学习与数据挖掘人才，是支撑国家人工智能战略和数字经济高质量发展的根基。本文提出的以创新能力培养为核心，通过重组课程框架，创新教学方法、拓展实践平台、重构评价机制的系统性教学改革方案，在实践中取得了积极成效，有效激发了学生的创新潜能，提升了其解决复杂实际问题的综合素养。改革贵在持续迭代，未来尤需顺应技术发展和社会需求，深化跨学科融合，并借助智能化手段，构建更加开放、协同、高效的创新人才培养生态，为人工智能时代输送更多拔尖创新人才。

参 考 文 献

- [1] 王静.机器学习算法在数据挖掘中的应用研究[J].软件,2024,45(09):115-117.
- [2] 李晋蓉,张高煜.机器学习实践课程教学改革[J].科技创新导报,2021,18(02):200-202.
- [3] 韦南,殷丽华,宁洪,等.本科“机器学习”课程教学改革初探[J].网络与信息安全学报,2022,8(04):182-189.
- [4] 王雷全,吴春雷,郭晓菲.机器学习科研实践课程建设[J].电子世界,2017,(17):50-51.
- [5] 董理,彭成斌,严迪群,等.基于层次化认知模型的本科机器学习课程教学改革[J].计算机教育,2023,(05):91-95.
- [6] 蒋磊,张丽,闫俊.面向本科教育的机器学习开放式实践课程教学改革探讨 [J].当代教育实践与教学研究,2020,(10):145-146+182.
- [7] 蔡赛华.数据分析与挖掘课程教学改革研究[J].电脑知识与技术,2022,18(31):178-180.
- [8] 曹付元,赵兴旺,高小方,等.一流课程建设背景下数据挖掘与机器学习课程教学改革 [J].计算机教育,2025,(07):155-159.
- [9] 杨书新,王振东,蔡虔,等.以科研能力为导向的“机器学习”教学改革[J].科技风,2024,(09):22-24.
- [10] 白露露,姚相宇.应用型本科机器学习课程教学改革探索与实践[J].河南财经学刊,2024,38(05):90-93.
- [11] 叶双,叶剑虹,雷庆,等.新工科背景下机器学习课程教学改革探索[J].福建电脑,2022,38(04):28-31.
- [12] 王重英.PBL模式下高校计算机专业课程思政教学创新研究——以“机器学习”为例[J].大学,2025,(15):97-100.