

# 校企合作理实双驱动下数据挖掘课程教学改革与研究

刘锋, 严芳

湖南工学院, 湖南 衡阳 421000

DOI: 10.61369/ETR.2025310033

**摘要:** 随着大语言模型 (LLMs) 与人工智能技术的快速发展, 数据挖掘课程教学面临理论与实践脱节、校企协同不足等核心挑战。本研究以“校企合作理实双驱动”为核心, 提出“理论-场景-工具”三维融合改革框架, 重构课程体系并创新教学模式。

**关键词:** 数据挖掘; 校企合作; 理实双驱动; 动态评价体系

## Teaching Reform and Research of Data Mining Course under the Dual Drive of School-Enterprise Cooperation and Integration of Theory and Practice

Liu Feng, Yan Fang

Hunan Institute Of Technology, Hengyang, Hunan 421000

**Abstract:** With the rapid development of Large Language Models (LLMs) and artificial intelligence technologies, the teaching of data mining courses is facing core challenges such as the disconnection between theory and practice, and insufficient school-enterprise collaboration. Focusing on the "dual drive of school-enterprise cooperation and integration of theory and practice", this study proposes a three-dimensional integration reform framework of "theory-scenario-tool", reconstructs the curriculum system and innovates the teaching mode.

**Keywords:** data mining; school-enterprise cooperation; dual drive of theory and practice; dynamic evaluation system

### 引言

随着大语言模型 (Large Language Models, LLMs) 技术的快速迭代与产业深度融合, 数据挖掘作为人工智能的核心技术之一, 正面临从“算法驱动”向“场景驱动”的范式转型。据清华大学经管学院《中国经济的数字化转型: 人才与就业》报告预测, 2025年我国大数据与人工智能领域人才缺口将突破200万, 其中兼具算法理解力与工程实践能力的复合型人才尤为紧缺<sup>[1]</sup>。然而, 当前高校数据挖掘课程普遍存在“重理论轻实践”“产教协同低效”等痛点: 一方面, 教学内容滞后于技术发展 (如大模型调优、隐私计算等前沿技术覆盖不足), 学生难以应对企业真实场景中的噪声数据处理、模型部署优化等挑战; 另一方面, 校企合作多停留于浅层实习对接, 企业因数据隐私风险、成本管控等因素参与动力不足, 导致教学与产业需求存在显著代际鸿沟<sup>[2]</sup>。

### 一、校企合作理实双驱动改革路径与实施策略

#### (一) 课程体系重构: 理论-场景-工具三维融合

##### 1. 模块化课程设计

将课程分为理论模块和实践模块。理论模块聚焦算法原理与数学基础, 结合企业需求优化教学内容。例如, 讲解梯度下降时, 同步引入电商销量预测场景, 推导损失函数优化过程。融入前沿技术 (如联邦学习、AutoML), 覆盖企业关注的隐私计算与

自动化建模<sup>[3]</sup>。

##### 2. 动态案例更新机制

联合企业每年更新10%的案例库, 纳入新技术 (如大语言模型调优、图神经网络)。例如, 与企业开发“医疗问答系统”, 集成 DeepSeek 的多轮对话能力, 覆盖文本生成与意图识别。

##### 3. 工具链集成: 开源+商用协同

例如, 使用 TipDM 平台对学生的校园消费数据进行采集、预处理和分析, 通过聚类分析、关联规则挖掘等算法对学生消费行

为进行分类、关联分析，构建消费预测模型<sup>[4]</sup>。

## （二）校企协同教学模式：轻量化合作与虚实结合

### 1. 校企合作共建虚拟仿真平台

虚拟仿真平台架构和功能包括数据脱敏层、沙箱训练环境和企业 API 接口。其中数据脱敏层利用差分隐私技术加密敏感数据（如用户身份证、医疗影像）。基于 Docker 容器化技术，集成脱敏数据集（如银行风控衍生库）与 AutoML 工具链。对接实时数据流（如电商订单量），模拟动态业务场景（如实时欺诈检测）<sup>[5]</sup>。

### 2. 双导师制与责任分工

聘请企业工程师担任企业导师，和校内导师共同承担课程教学。校内导师理论教学主导者与课程设计协调者，讲解数据挖掘算法原理（如决策树、聚类分析）、数学基础（微积分、概率论），结合企业需求优化教学大纲。企业导师是实践能力培养者与行业经验传递者，提供企业真实场景案例（如金融风险模型调优），指导学生完成数据清洗、模型部署全流程，评估学生项目成果的落地价值（如模型预测误差是否达标），提出改进建议<sup>[6]</sup>。

### 3. 分层实践设计

分层实践设计遵循“基础技能→综合应用→创新突破”的递进逻辑，结合校企合作资源，构建从工具掌握到业务落地的完整能力链。

基础实验的目标是让学生掌握数据挖掘核心工具与基础算法，解决标准化问题。其内容包括：

（1）数据挖掘工具学习：如 TipDM、Python 等，掌握这些工具的基本操作和使用方法。

（2）数据分析基础：包括数据清洗、数据可视化、数据转换等基础技能，让学生能够对原始数据进行预处理，使其适合进行进一步的分析。

（3）机器学习基础算法：如决策树、聚类算法、关联规则挖掘等，理解这些算法的基本原理，并能够应用这些算法解决简单的数据挖掘问题。

实践项目：设置一些简单的练习和实验，让学生在实际操作中巩固所学的基础知识和技能。例如，使用 TipDM 进行简单的数据导入、清洗和可视化操作；利用 Python 的数据挖掘库（如 Scikit-learn）实现决策树分类器等<sup>[7]</sup>。

综合应用项目的目标是培养学生的综合应用能力，使其能够将多种技能和知识结合起来，解决较为复杂的数据挖掘问题。内容包括：

（1）实际案例分析：结合企业提供的实际业务数据，分析业务需求，设计数据挖掘方案。例如，分析电商销售数据，挖掘客户的购买模式和偏好。

（2）特征工程与模型优化：学习如何进行特征选择、特征提取和特征构造，以提高模型的性能。同时，掌握模型优化的方法，如参数调整、交叉验证等。

（3）数据挖掘项目全流程实践：包括数据预处理、模型构建、模型评估、结果分析等环节，让学生熟悉数据挖掘项目的完整流程。

实践项目：开展综合性的数据挖掘项目，要求学生运用所学的多种技能和知识，完成从数据准备到模型部署的全过程。例如，企业提供的客户流失预测项目，学生需要收集和整理客户数据，选择合适的算法构建预测模型，并对模型进行评估和优化。

创新突破与前沿探索项目的目标是激发学生的创新思维，培养其在数据挖掘领域的前沿探索能力和创新能力。内容包括：

（1）前沿技术研究：学习和探索数据挖掘领域的前沿技术，如深度学习、图神经网络、联邦学习等在数据挖掘中的应用。

（2）创新项目开发：结合企业的实际业务需求，开展创新性的数据挖掘项目。例如，利用深度学习技术优化图像识别模型，提高图像分类的准确性<sup>[8]</sup>。

（3）科研课题与企业横向项目参与：鼓励学生参与教师的科研课题或企业的横向项目，与团队成员合作解决实际问题，在实践中提升创新能力。

实践项目：鼓励学生自主设计和开发创新性的数据挖掘项目，或者参与科研课题和企业项目。例如，开展基于区块链技术的数据挖掘安全性和隐私保护研究等<sup>[9]</sup>。

## （三）教学资源共建：数据、工具、案例一体化

### 1. 企业案例库开发

提炼企业真实问题（如金融反欺诈、社交网络分析），设计全流程任务：

数据清洗（处理缺失值、异常值）→特征工程（IV 值筛选、Embedding 生成）→模型部署（Flask API 封装）。

案例：某银行信用卡风控项目，学生利用 XGBoost 构建模型，将误报率从 15% 降至 8%。

### 2. 合成数据与隐私保护

利用 GAN 生成脱敏数据（如 StyleGAN2 生成医疗影像），特征保留率  $\geq 90\%$ 。

示例：某合作医院提供脱敏 CT 数据，学生训练肺结节检测模型（AUC=0.89），接近真实数据效果（AUC=0.91）。

### 3. 竞赛资源转化

将 Kaggle、天池竞赛、数据挖掘竞赛任务融入课程（如“电商销量预测”“5G 基站能耗优化”）。

动态评价体系：过程量化与结果导向

#### 1. 动态评价体系的构建原则

过程量化：尽可能采用可量化的指标，减少模糊评价，提高准确性。利用大数据和人工智能等技术，对教学过程进行全周期、多角度、实时动态的量化评价。

结果导向：评价内容应与教学目标紧密相连，确保评价结果能够真实反映学生的学习成果和教师的教学效果，强调评价结果的实际应用，促进学生和教师的共同成长。

持续反馈：建立有效的反馈机制，鼓励双向沟通，促进学生和教师及时了解自己的学习和工作状态，发现不足并进行改进。

多元主体：评价主体应多元化，包括教师、学生、企业专家、行业人士等，通过多角度的评价，全面了解学生的学习情况和教师的教学表现<sup>[10]</sup>。

#### 2. 动态评价体系的具体实施方法

##### （1）过程量化

建立过程性评价指标体系：

学习过程指标：包括出勤率、课堂表现、作业完成情况、小组讨论参与度等，通过学习管理系统（LMS）等工具进行数据采

集和分析。

**技能掌握指标：**通过在线测试、实验操作、项目实践等方式，及时了解学生对知识和技能的掌握情况。

**学习进度指标：**根据教学计划和学生的学习任务完成情况，评估学生的学习进度是否符合预期。

采用多维度评价方式：

**自评与互评：**引导学生进行自我评价和同学之间的互评，帮助学生了解自己的学习情况和存在的问题，同时培养学生的评价能力和团队合作精神。

**教师评价：**教师根据学生的课堂表现、作业、考试等综合情况进行评价，及时发现问题并给予指导。

**企业评价：**企业导师对学生的实践项目、实习表现等进行评价，提供真实的工作场景和行业标准的反馈。

## (2) 结果导向

明确评价目标与标准：

根据教学目标设定评价目标：评价目标应与课程目标和教学计划相一致，明确学生在知识、技能、态度等方面应达到的水平。

制定科学合理的评价标准：评价标准应具体、明确、可操作，能够准确衡量学生的学习成果和教师的教学效果。

注重评价结果的应用：

与成绩评定挂钩：评价结果应作为学生成绩评定的重要依

据，合理分配过程性评价和终结性评价的权重，全面反映学生的学习情况。

**与教学改进挂钩：**教师应根据评价结果及时调整教学方法和策略，改进教学内容和教学过程，提高教学质量。

**与职业发展挂钩：**将评价结果与学生的学业发展和职业规划相结合，为学生提供个性化的学习建议和职业指导。

## 二、教学改革实施情况与成效

我院在2022年、2023年和2024年分别对2020级、2021级和2022级应用统计学专业的本科生实施“数据挖掘”课程创新教学改革。通过学习管理系统记录学生的出勤、作业、在线测试等情况，每周发布学习报告；企业导师对学生的实践项目进行中期和期末评价，给出详细反馈。

## 三、结论

本文通过课程重构实现理论、场景与工具融合，借助轻量化校企协同教学打破实践瓶颈，以分层实践设计助力学生能力进阶，推动教学资源共建支撑多元教学活动，以动态评价体系保障教学质量与学生发展。

## 参考文献

- [1] 清华大学经济管理学院. 中国经济的数字化转型：人才与就业 [R]. 北京：清华大学，2020.
- [2] 国务院. “十四五”数字经济发展规划 [Z]. 北京：中国政府网，2021.
- [3] 智联招聘. 人工智能与大数据人才需求洞察报告 [R]. 北京：智联招聘研究院，2024.
- [4] 王维，孔超. 基于问题导向的混合式数据挖掘课程教学模式探索 [J]. 电脑知识与技术，2025，21(4)：170-173.
- [5] 吴飞，邓哲颖，肖婉. 基于多模态学习的数据挖掘课程教学设计与实践 [J]. 集成电路应用，2024，41(12)：406-408.
- [6] 马莹莹，崔文昊. 数据科学专业核心课程的教学设计与思考——以统计学习课程为例 [J]. 高教学刊，2024，10(33)：42-45.
- [7] 赵瑞晓，顾航，孟秋云. 基于“科教融合、以赛促学”的高职院校大数据挖掘与分析课程教学改革研究 [J]. 信息与电脑（理论版），2024，36(21)：217-220.
- [8] 赵晓伊，王莎莎. 项目主线法教学在数据挖掘课程的改革与实践 [J]. 电脑与电信，2024，(11)：92-96.
- [9] 任和，李萍，张立方，等. 基于知识元融合的健康数据挖掘课程设计探索 [J]. 中国多媒体与网络教学学报（上旬刊），2024，(11)：68-71.
- [10] 张金芳，李亚男，贾冬青，等. 数据挖掘课程的教学与评价模式改革研究 [J]. 信息与电脑（理论版），2024，36(19)：254-256.