Education Reform and Development, 2026, Volume 8, Issue 5
https://ojs.bbwpublisher.com/index.php/ERD
Online ISSN: 2652-5372
Print ISSN: 2652-5364

Exploration and Practice of Generative AI-Embedded Teaching Paths for Data Mining Courses

Lulu He*, **Yue Huang**

Anhui Polytechnic University, Wuhu 241000, Anhui, China

Author to whom correspondence should be addressed.

Abstract: In the digital era of rapid artificial intelligence development, there is an urgent demand in the data mining field for talents with integrated "algorithm + business" capabilities. Addressing issues such as high thresholds for code implementation, shallow understanding of algorithm principles, and difficulties in implementing optimization strategies in traditional data mining teaching, this study focuses on cultivating the data mining capabilities of students in computer majors. Taking the Python ecosystem as the core teaching tool, generative AI tools are organically integrated as auxiliary means into the entire teaching process. Relying on the "Enterprise Customer Churn Prediction" project, the paper designs specific intervention points of AI tools in key links such as data exploration, model construction, algorithm optimization, and result interpretation. Practice shows that this model can effectively reduce programming cognitive load, stimulate students to explore algorithm optimization logic, and improve the quality of model implementation, thereby addressing students' fear of difficulties, enhancing their comprehensive capabilities to solve complex engineering problems, and providing a practical and referable path for digital teaching reform in computer courses in similar institutions.

Keywords: Computer education; Data mining; Generative AI; Embedded auxiliary teaching; Project-based learning

Online publication: June 30, 2026

1. Introduction

The global industrial digital transformation has driven a shift in demand for data science talent toward in-depth analysis and value creation. Currently, university data mining courses face challenges in competency development: traditional instruction focuses heavily on algorithmic formula derivation or static code demonstrations, neglecting the practical processes of interactive analysis and model tuning. Generative AI technologies have not been effectively integrated, making it difficult for students to use intelligent tools to improve coding efficiency and analytical depth. During teaching, students with weak foundations often develop learning anxiety, and the conventional "demonstration–imitation" model results in insufficient ability to independently solve real-world data problems.

In recent years, the application of generative artificial intelligence (Generative AI) in education has become an international research hotspot. A 2023 guidance document by UNESCO emphasizes that AI should serve as a tool to augment human capabilities rather than replace them, with a focus on ethical human–machine collaboration and competency development [1]. A study by Kasneci et al. (2023) in Learning and Individual Differences indicates that large language models (LLMs) have significant potential in personalized tutoring and reducing cognitive load, yet also pose challenges related to academic integrity and intellectual inertia [2]. For computer-related majors, Prather et al. (2023) proposed at the ACM SIGCSE conference that generative AI is reshaping computing education, requiring instructors to shift their role from knowledge transmitters to designers and facilitators of the learning process [3]. Against the backdrop of domestic educational digitalization, Zhu et al. (2022) elaborated the practical logic of educational digital transformation [4], arguing that technology-enabled education must return to the essence of nurturing talent, providing theoretical support for the deep integration of AI and curricula [5]. Xiao et al. (2023) further constructed a scenario framework for generative AI-enabled online learning, verifying AI's effectiveness in personalized learning path recommendation and interactive support [6]. Regarding programming assistance, Dakhel et al. (2023) empirically examined the impact of tools such as GitHub Copilot on code quality, finding that they effectively improve development efficiency while requiring users to possess adequate code review capabilities [7].

Furthermore, Zhang et al. (2026) explored AI-assisted approaches in financial data visualization teaching, confirming that embedding AI as an auxiliary tool can effectively lower learning barriers and enhance analytical quality [8]. Inspired by this, this study focuses on exploring the organic integration of accessible generative AI tools (e.g., DeepSeek, Copilot, etc.) into data mining instruction and designing a practical teaching model. It aims to use AI as a catalyst to empower the entire teaching process, systematically improve students' data mining and algorithm optimization abilities, and offer operable practical references for the digital teaching reform of computer-related majors.

2 Core concept: AI tools as "embedded" assistants in data mining teaching

To effectively integrate AI tools into data mining teaching, their positioning must first be clarified. This study argues that AI should not be treated as an independent tool parallel to Python programming that demands additional class hours; instead, it should act as an intelligent tutor deeply embedded within the standard learning path of data mining. Its core value lies in serving as a learning companion, providing instant, low-threshold support when students encounter cognitive obstacles, thinking bottlenecks, or expressive difficulties. This flattens the steepness of the overall learning curve, frees up students' cognitive resources, and allows them to focus on more core tasks such as constructing algorithmic logic and training higher-order thinking. Based on this positioning, an assistance framework for AI tools across key teaching stages of data mining is constructed (**Table 1**). Taking the natural progression of a data mining project as the main line, the framework defines the specific roles and typical application scenarios of AI at each critical stage, ensuring targeted and effective integration.

Table 1. Auxiliary positioning and application scenarios of AI tools in various stages of data mining teaching

Core Teaching Stage	AI Tool Auxiliary Positioning	Key Competency Training Objectives	Typical Application Scenarios
Stage 1: Data Preparation & Exploration	"Data Dictionary & Cleaning Consultant"	Understand data structures, complete basic data acquisition and cleaning.	Input a list of fields from an unfamiliar dataset and ask AI to explain their business meanings; when encountering cleaning errors (e.g., "data type mismatch"), seek troubleshooting ideas and solutions from AI.
Stage 2: Model Construction & Code Implementation	"Code Logic Interpreter & Example Library"	Understand algorithm principles, master core code writing and debugging.	Paste complex algorithm code into AI and request "explain the calculation logic of this code in plain language"; describe business calculation requirements (e.g., "construct feature crosses") to AI to obtain an initial code draft for learning and modification.
Stage3: Algorithm Optimization & Tuning	"Optimization Strategy Enthusiast"	Master hyperparameter tuning logic and improve model performance.	When model performance is poor, ask AI, "Besides adjusting the learning rate, what other optimization strategies can be tried?"; use AI to analyze the causes of anomalies in the Loss curve and get improvement suggestions.
Stage 4: Result Interpretation & Report Writing	"Analysis Inspiration Trigger & Report Polishing Assistant"	Select appropriate evaluation indicators and organize rigorous analysis reports.	Input scattered data conclusions and ask AI to organize them into clearly structured, standard analytical paragraphs; use AI to generate a first draft of SHAP value explanations, which students then revise professionally.

The establishment of this framework ensures that the application of AI tools is no longer arbitrary or superficial, but purposeful and structured in serving the achievement of core data mining competency objectives, realizing the resonance between auxiliary tools and mainline teaching.

3. Overall instructional design

To put the above concepts into practice, this study designs an "AI-integrated" teaching path underpinned by a complete data mining project running through the entire course. Taking Project-Based Learning (PBL) as the carrier, AI-assisted tasks are integrated as a natural component of project progression [9,10].

3.1. Project initiation and data exploration (AI-assisted comprehension and framework construction)

At the beginning of the course, students work in groups to receive assigned tasks, and the instructor provides desensitized enterprise datasets. Before direct manipulation, students first conduct data reconnaissance: they input a list of field names into an AI tool to inquire about their business meanings and correlations. With explanations generated by AI, students can quickly grasp the business context of the data, laying a foundation for analysis and effectively avoiding aimlessness and confusion when confronted with raw data.

3.2. Model construction and code implementation (AI-assisted learning and difficulty breakthroughs)

Establishing a proper data processing pipeline in Python is fundamental to analysis, while writing algorithmic code is critical to implementing business logic, both of which are common pain points for students. At this stage, the instructor's role shifts from "the sole provider of code" to "a logic interpreter and debugging facilitator". When students encounter obstacles in algorithm implementation, they are encouraged to describe their computational intentions in Chinese to AI. After obtaining code suggestions, students' core task is to test, understand the logic line by line, and debug errors in the runtime environment. This process transforms code learning from "memorizing syntax" to "verifying logic and solving problems", significantly improving learning autonomy [11].

3.3. Algorithm optimization and interactive analysis (AI-assisted thinking and insight discovery)

Students are required to compare the performance of different models through experiments and design optimization strategies. When confused about parameter tuning directions or analytical perspectives, AI can act as a design consultant and recommend appropriate optimization combinations. More importantly, when students identify model overfitting or underfitting, they can immediately ask AI about potential technical causes. Targeted hypotheses provided by AI, such as "insufficient regularization strength" and "curse of dimensionality", guide students in conducting focused technical drill-down analyses, systematically cultivating their in-depth analytical thinking of problem identification → hypothesis formulation → verification and confirmation [12].

3.4. Analysis report generation and result presentation (AI-assisted expression and outcome enhancement)

The ultimate value of analysis lies in clear communication. To address students' inadequate ability in writing technical reports, AI can serve as a first-draft writing partner. Students organize core conclusions, key indicators, and model performance metrics into a list and input it into AI, instructing it to generate structurally rigorous and professionally phrased core sections of the report. However, the key to teaching lies in the subsequent critical revision: students must review, revise, supplement, and deepen AI-generated content by combining specific details from dashboards, unique group insights, and learned algorithm theories. This process not only improves report quality but also trains students' comprehensive professional literacy in information evaluation, perspective integration, and precise expression.

4. Teaching implementation process

To illustrate the above teaching path in detail, a case study of a comprehensive project implemented in the "Data Mining" course is presented below.

4.1. Project background and core tasks

Students are provided with five years of user behavior data from a telecommunications company. The core task is to build a dynamic prediction model using Python, diagnose the causes of user churn, conduct in-depth analysis of driving factors and potential risks, and finally produce a concise analysis report and a runnable code repository.

4.2. Project implementation process

The project is implemented over 4 class sessions (8 class hours):

(1) First session: Analysis, Planning, and Data Exploration

Each group first uses AI tools for brainstorming. They input the prompt: "As a data scientist, to conduct a comprehensive diagnosis of user churn in a telecommunications company, what core dimensions should be analyzed? What key features can be calculated under each dimension?" The systematic analysis framework provided by AI effectively helps groups move beyond isolated data discussion and develop more professional analysis plans.

(2) Second to third sessions: Tool Implementation and In-depth Analysis

Many groups encountered difficulties when constructing the complex feature "weighted user activity score". They described to AI: "I have a 'user table' and a 'call detail table'; I want to calculate the weighted activity level for each user." After obtaining the initial code, they debugged and understood the contextual logic under the guidance of the instructor. During the optimization stage, one group found that the model performed poorly on the test set. They immediately asked AI: "Besides increasing data volume, what technical factors may cause poor generalization ability of a classification model?" Concepts such as data leakage and class imbalance in AI's response inspired the group to further analyze data distribution. The problem was finally solved using SMOTE sampling, demonstrating deeper insights [13].

(3) Fourth session: Report Integration and Presentation

After organizing analytical findings, each group used AI to assist in generating the first draft of the report. For example, they input key points: "Overall churn rate increased by 5%; mainly due to declining satisfaction among high-tariff package users; it is recommended to optimize package structure." and requested coherent conclusions. After AI generated the text, the group enriched the content and strengthened logical reasoning by incorporating specific confusion matrix data and finer-grained causal analysis, forming the final report.

4.3. Project implementation effects

This case shows that the integration of AI tools has successfully shifted the focus of learning from "how to write code" to "how to think about data problems". AI plays an irreplaceable auxiliary role in three aspects:

(1) First, as a thinking expander, it provides an analysis framework in the early stage of the project.

(2) Second, as a difficulty decoder, it offers solutions when technical bottlenecks occur.

(3) Third, as an outcome amplifier, it improves the standardization of written outputs.

The whole process reflects the principle of "human–machine collaboration, people-oriented", with students' core abilities of analysis, judgment and decision-making always in the dominant position [14].

5. Teaching evaluation and reflection

To comprehensively evaluate the effect of the "Data Mining + AI" integrated teaching model, this study designed a diversified evaluation system focusing on process and comprehensive abilities.

5.1. Teaching evaluation

The evaluation mainly includes three dimensions:

(1) Code and technical outcomes (50%): assessing data preprocessing, model construction and optimization effects.

(2) AI tool application process (30%): through submitted "AI usage logs", evaluating students' accuracy in posing questions, initiative in usage, and ability to critically process outputs.

(3) Comprehensive business analysis ability (20%): assessing business insight and logical reasoning through reports and presentations.

5.2. Teaching outcomes

First, the classroom atmosphere became more active; students' frustration with technical difficulties decreased, and behaviors of active exploration and solution-seeking increased significantly. Second, the overall quality of final project outcomes improved, with more robust models and markedly better structural integrity and linguistic professionalism in analysis reports. Most importantly, student feedback indicated that this teaching method made them "feel that they truly used data to solve real-world problems", transforming their perception of data mining work from abstract to concrete.

5.3. Challenges and improvements

In practice, this study also encountered challenges and explored corresponding strategies:

(1) Challenge 1: Over-reliance on AI and intellectual inertia

Some students attempted to use AI to generate all the code directly.

Strategy: Assign high weights to "AI usage logs" and "personal revision notes" in the evaluation criteria, and conduct special discussions on "limitations of AI-generated code" to strengthen awareness of critical verification.

(2) Challenge 2: Uncontrollability of AI output

AI may provide incorrect code or generic explanations.

Strategy: Turn this into a valuable teaching opportunity. Instructors guide students to treat erroneous AI outputs as "debugging and error correction" exercises, analyzing causes together to achieve a deeper understanding of correct logic.

(3) Challenge 3: New requirements for teacher competence

Teachers need to accelerate their role transformation from "one-way transmitters" to "learning process designers".

Strategy: The teaching research team conducts collective lesson preparation, jointly develops a series of "AI-assisted learning task sheets", shares high-quality AI prompt cases, and focuses on designing in-class interactive activities that cannot be replaced by AI, enhancing interpersonal interaction and high-level thinking collision.

6. Conclusion and outlook

Against the urgent demand for computer talent cultivation driven by digital transformation, this study, grounded in actual teaching practice in universities, explores a new competency development pathway that uses generative AI tools as a flexible and intelligent auxiliary line, focusing on application scenarios in the data mining course. Practice has proven that this pathway can effectively leverage the unique advantages of AI in lowering the initial technical threshold, providing instant learning support, inspiring multi-dimensional analytical thinking, and improving the quality of outcome presentation. Consequently, it helps students focus their learning on business logic construction and complex problem-solving, effectively enhancing their comprehensive practical capabilities in data mining and analysis.

This study offers a transformation in teaching thinking oriented toward the digital future: educators should proactively embrace changes, make good use of intelligent technologies as a lever to empower teaching, and unleash students' creativity. Looking ahead, this model can be further deepened in the following aspects: first, developing an "AI-assisted learning resource package" that precisely matches the core knowledge modules of data mining; second, exploring the establishment of a long-term tracking and evaluation model for the development of students' "human–machine collaboration" capabilities; third, extending this integrated concept to other professional courses and gradually building a curriculum system supporting digital skill development, to systematically cultivate interdisciplinary technical talents who understand business, master tools, and are capable of effective collaboration.

Funding

College Students' Innovation and Entrepreneurship Training Program of Anhui Polytechnic University (Project No.: S202510363363, S202510363339); Scientific Research Startup Foundation for Talents of Anhui Polytechnic University (Project No.: S022023033, S022024053); National Natural Science Foundation of China (Project No.: 62401011); Natural Science Foundation of Anhui Province (Project No.: 2408085QF202, 2508055MD158); Scientific Research Project of the Education Department of Anhui Province (Project No.: 2025AHGXZK20078); Scientific Research Startup Foundation for Talents of Tongling University (Project No.: 2024tlxyrc025); High-Quality Course of Ideological and Political Theories in Teaching—Financial Data Mining, Quality Engineering Project of Anhui Polytechnic University; Quality Engineering Project of Anhui Polytechnic University (Project No.: 2024jyxm38)

Disclosure statement

The authors declare no conflict of interest.

References

[1] UNESCO, 2023, Guidance for Generative AI in Education and Research. UNESCO, Paris.

[2] Kasneci E, Sessier K, Küchemann S, et al., 2023, ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learning and Individual Differences, 103: 102274.

[3] Prather J, Denny P, Leinonen J, et al., 2023, The Robots Are Here: Navigating the Generative AI Revolution in Computing Education. Proceedings of the 54th ACM Technical Symposium on Computer Science Education, 2023: 1085–1086.

[4] Chan CKY, 2023, A Comprehensive AI Policy Education Framework for University Teaching and Learning. International Journal of Educational Technology in Higher Education, 20(1): 38.

[5] Zhu Z, Hu J, 2022, Practical Logic and Development Opportunities of Educational Digital Transformation. E-Education Research, 43(1): 5–15.

[6] Xiao J, Bai Q, Chen M, Lu L, 2023, Generative AI Empowers Online Learning Scenarios and Implementation Pathways. E-Education Research, 44(9): 12–20.

[7] Dakhel A, Majdinasab V, Nikanjam A, et al., 2023, GitHub Copilot AI Pair Programmer: Asset or Liability? Journal of Systems and Software, 203: 111734.

[8] Zhang M, Zhang Y, Wen Y, 2026, Exploration and Practice of AI-Assisted Teaching Paths for Financial Data Visualization and Analysis. Education Reform and Development, 8(1): 282–288.

[9] Chen X, Zhou J, Liu H, 2025, Research and Practice on Teaching Reform of Data Mining Course for Emerging Engineering Education. The Theory and Practice of Innovation and Entrepreneurship, 8(23): 47–49.

[10] Wang Z, Hu X, Li L, et al., 2025, Research on Teaching Content and Mode Reform of Data Mining Courses Assisted by Large Language Models. Computer Education, (11): 74–79.

[11] Tang Y, Jiang Y, Xiong L, et al., 2025, Intelligent Teaching Assistant for Data Mining Course Based on RAG. Computer Technology and Development, 35(12): 141–148 + 182.

[12] Huang J, Meng D, Wu X, 2025, Teaching Reform and Practice of Data Mining Technology Course Driven by Integration of Science and Education. Journal of Anhui University of Technology (Social Sciences), 42(03): 50–52.

[13] Chen C, Xu H, Liu P, 2025, Teaching Reform and Practice of Data Mining Course Empowered by Artificial Intelligence. Computer Knowledge and Technology, 21(14): 138–140.

[14] Gao X, Cao F, Zhao X, et al., 2025, Exploration of Personalized Teaching Mode of Data Mining Course under the Background of Emerging Engineering Education. Western China Quality Education, 11(06): 149–153.