Journal of Contemporary Educational Research, 2026, Volume 10, Issue 5
http://ojs.bbwpublisher.com/index.php/JCER
Online ISSN: 2208-8474
Print ISSN: 2208-8466

Research on the Application Effect of Project-Based Learning (PBL) in the Practical Teaching of Big Data Mining

Yangping Tong

Hubei Business College, Wuhan 430079, Hubei, China

Abstract: With the rapid development of the big data industry, the demand for data mining talents in the market has significantly increased, and higher requirements have been imposed on the practical abilities of these talents. This has compelled schools to accelerate the innovation of teaching methods. Project-based learning has obvious application value in the practical teaching of big data mining. Based on this, this paper takes the textbook *Python Data Mining Algorithms and Application Experiments and Course Training Guidance* as the core carrier, explores the application path of project-based learning in the practical teaching of big data mining, evaluates the effect of teaching application, builds a systematic teaching system, and designs targeted teaching content and teaching plans. The research results can enhance students' practical operation ability and teamwork ability, and provide certain references and inspirations for promoting teaching reform.

Keywords: Big data; Data mining; Project-based learning

Online publication: June 9, 2026

1. Introduction

In the era of digital economy, big data technology has become the core driving force to promote industrial upgrading, and the demand for data mining professionals continues to grow. Data mining, as a highly practical interdisciplinary field, requires students to not only master theoretical knowledge but also possess the ability to solve actual business problems. Project-based learning (PBL) is a student-centered approach that enables students to work through a project in a variety of real-world scenarios and to use, build, and improve their problem-solving skills within the project. Based on the textbook system of *Python Data Mining Algorithms and Application Experiments and Course Training Guidance* published by Tsinghua University Press in 2024, this paper systematically analyzes the application scheme of project-based learning method, so as to provide some references for the actual teaching of big data mining in universities [1].

2. Value analysis of project-based learning application in big data mining teaching

PBL teaching guides students to complete projects through real and complex project tasks, guides them to carry out independent exploration and team cooperation, so as to master relevant knowledge and skills, and cultivate comprehensive ability. In the teaching process, teachers create real driving questions for students, allowing them to continuously explore the essence of PBL during the project completion process. Teaching is student-centered, and teachers need to provide students with diverse task resources and organize them to collaborate in the process of completing the project, and evaluate the comprehensive project outcomes of students. Currently, a complete PBL learning process is divided into five stages: initiation, planning, execution, monitoring, and conclusion. During this process, teachers no longer play the role of the dominant figure but instead act as guides, supporters, and evaluators. The learning of knowledge is driven by the students themselves, and the process of mutual collaboration is where the knowledge is accomplished, which is not directly taught by the teacher.

Engineering practice teaching aims to cultivate applied talents. The engineering practice ability, innovative thinking, and problem-solving ability of students are of vital importance, and this is consistent with the talent cultivation direction of PBL. The big data mining course is a typical engineering practice teaching course, and its knowledge system includes data collection, preprocessing, modeling, evaluation, and deployment, etc. These links have obvious practicality and require students to be able to comprehensively apply multidisciplinary knowledge to solve practical business problems. Therefore, the application of PBL teaching in big data mining practice can improve students' mastery of the entire knowledge of big data mining through project-based task design, enhance students' motivation for knowledge learning and knowledge transfer ability through real business scenarios, and enable students to exercise communication skills and professional qualities in the process of completing project tasks, thereby improving the targeted and effectiveness of talent cultivation [2].

3. Design of PBL teaching scheme based on textbooks

3.1. PBL design for the basic experiment module

The basic experiment module helps students master the basic tools, core algorithms, and basic operation procedures of data mining. Each experiment is designed as an independent small-scale PBL project, with a class duration of 2–4 hours. For the experimental projects related to the configuration of Python development environment and common database processing in the textbook, PBL teaching focuses on the analysis of small datasets, and specific task types are carried out. Students are required to independently complete the environment setup, data loading, basic statistical analysis, and data visualization output. They are encouraged to apply different tools to complete the project tasks. For data collection, data cleaning, feature engineering, and data visualization, PBL teaching adopts the method of introducing dirty data in real scenarios as the project task, allowing students to independently identify the quality issues of dirty data, select appropriate data preprocessing methods, output standard data sets that meet the requirements of PBL teaching, and students can understand data cleaning methods, master this ability, and establish the thinking of feature engineering during the process of completing the project. This thinking cultivation is also the most core part of data mining projects. How to form effective feature engineering thinking and implement it is related to the final quality of talent cultivation. For the projects related to classification algorithms such as decision trees, naive Bayes, neural networks analysis, and core algorithms such as clustering algorithms, association rules

algorithms in the textbook, PBL teaching requires dividing the algorithm experiments into multiple types of comparisons.

Students were required to apply different algorithms to achieve the goals of the project and to compare the actual effects of different parameter Settings and different models. Analyzing the advantages and disadvantages of algorithms in different application scenarios will enable them to be more proficient in choosing the target algorithm. This deepens their understanding of the algorithm principles. For model evaluation experiments such as classification evaluation, clustering quality evaluation, and model optimization in textbooks, PBL teaching requires students to select an appropriate index system for the completed modeling task and form the final result of index evaluation. The self-built index model is also used to verify the effect of the solution. Through this project task, students' model evaluation ability and optimization thinking can be effectively exercised, thus ensuring the quality of the data project.

3.2. Integrated practical training module PBL design

3.2.1. Problem-oriented teaching

The integrated practical training module of PBL should emphasize problem-driven and strengthen the integration of knowledge points. Taking the practical training project "Predicting House Prices for Second-hand Houses" in the textbook as an example, the teacher should break the isolated situation of knowledge points and connect them together. Students should be guided to actively connect relevant knowledge points such as data collection and missing value handling based on the goal of "How do real estate agents choose scientific pricing tools for second-hand houses?" For better data collection, the students' projects should also include regression model training and model interpretation, etc. These cross-chapter knowledge points can all be integrated into the design of a single project. In teaching, the teacher should not specify any specific algorithms but allow the student teams to sort out the specific factors that can meet the customer's needs in the business scenario. For example, if a customer wants to choose a house within the target psychological price range, then the tool should provide the function of "feature screening." To better collect data, the students' projects should also include regression model training, model interpretation, etc. These cross-chapter knowledge points can all be integrated into the design of a single project. In this process, students not only need to output the model prediction results but also explain the weight of different factors on the house price [3]. Finally, as a real estate agent, they should provide customers with a visual pricing tool. This exploration and output process enables students to fully understand why they choose this parameter, why they explain the model in this way, and thus actively integrate knowledge to complete the application.

3.2.2. Outcome generation and value implementation

The core logic of the PBL learning method lies in setting the outcome as the ultimate goal, allowing students to fully experience the process of the project. The project outcomes of students must possess the potential for practical implementation. Teachers should guide students to fully consider whether the business value can be implemented during the process of completing the data mining learning. Students should develop a business mindset and be responsible for their own projects. Using the case of supermarket product sales in the textbook as an example, some students thought that as long as they output the rule list, they had completed the task. However, teachers should guide students to use outcome orientation and implementation effect as the measurement goals of the project, requiring students to transform into implementable business and organize students to conduct thorough research on the display of goods in physical supermarkets, understand the logic

of the displayers, and analyze the correlation between product display and promotional effects. Focusing on the current promotion activities and the real needs of consumers, the algorithm in the textbook is used to mine the transaction data of the supermarket, to understand the correlation rules between product display and promotional effects. During this process, students will break out of the fixed thinking pattern of rule output. By filtering out some invalid rules based on the business scenario and mining out the display logic more relevant to user behavior, for example, users who purchase outdoor barbecues are more likely to repurchase charcoal; users who purchase milk powder also pay more attention to children's vitamins. Through the mining of hidden correlation rules, a more practical display plan will be output, helping the stockers of the supermarket to place related products adjacent to each other and combine pricing. During this process, the data association output by students can provide support for the product display of the supermarket, and the quantitative tool of high-value rules formed by students can truly support the increase in sales. During this process, students can also truly feel that the application of algorithms has practical commercial value. The algorithm can ultimately solve problems in business, and with this application thinking, students will be more interested in learning data mining skills.

3.2.3. Scenario-based data mining teaching

Data mining is highly targeted towards specific scenarios. When conducting PBL teaching, teachers should use scenarios as the starting point to train students' ability to mine and analyze data in specific scenarios, rather than merely mastering some algorithms. Teachers should design more practical scenarios to avoid deviations in the application of technology, which could lead students to model for the sake of modeling. Taking the practical training project of precise marketing for bank financial products in the textbook as an example, when designing PBL project tasks, teachers should introduce the pain points of real bank marketing, and require students to conduct data sorting on the current situation of wasted bank marketing resources. Through the entire process of data analysis, based on the feature database provided in the textbook such as customer age, income, purchase records, risk preference, etc., predict the purchase intentions of customers, and select the data sorting model. During the process of students choosing the model to apply, teachers should guide students to think about the evaluation deviations generated by different models in specific business scenarios. The selection of models should not only pursue overall accuracy, but also focus on the recall rate of high-potential customers and the effective mining of long-tail users, that is, through the model, determine whether some customers have purchasing potential, cover all customers with real purchase intentions, and reduce the churn of high-value customers. Through this scenario-based thinking, students can apply models to output customer stratification lists more accurately, select differentiated marketing phrases, and imitate the specific decision-making process of real bank marketing teams. Teachers should provide students with more data support systems to help them master the application of specific data mining algorithms such as logistic regression and decision tree classification, and understand that algorithm indicators should serve the core of technical learning, which is to serve business development, forming a thinking mode of integrating technology and business.

3.2.4. Ability-based tiered teaching

The learning of data mining and application algorithms, as well as the learning of decision-making thinking, have high requirements for students' comprehensive abilities. Students at different levels have significant differences in the development of their thinking and the mastery of their abilities. In specific teaching,

teachers should focus on the tiered task goals of PBL teaching and design differentiated content levels. It is necessary to meet the expansion needs of students with strong foundations and the basic learning needs of students with learning difficulties. In the setting of PBL teaching goals, teachers can divide project tasks into three difficulty levels: the basic level requires all students to use the K-Means algorithm in the textbook to complete the calculation of customer RFM characteristics and achieve basic customer value segmentation; the intermediate level requires students with sufficient ability to try different algorithms such as DBSCAN and hierarchical clustering, and compare the differences in the effects of different clustering methods; the innovative level requires students to combine user browsing behavior, social attributes, and other extended features to construct more detailed customer profiles. The final requirement of the project is that each team outputs differentiated operation strategies: for high-value customers, design exclusive member benefits; for high-risk churn customers, design recall activities; for new customers, design first-purchase incentive plans. In the specific result assessment, teachers can use the clustering evaluation indicators in the textbook as the assessment standard for students at the basic level. For the intermediate level, students' strategy choices should have certain innovativeness and provide feasibility reports. For the innovative level students, they should have combinability in strategy selection and provide actual effect analysis. Through this differentiated teaching based on individualized instruction, a more practical task logic is created, and the adaptability of PBL teaching is enhanced.

4. PBL teaching implementation guarantee system

PBL teaching requires high standards for the institutional resources provided by the school and the process control requirements. Schools and teachers should provide more guidance to students and offer sufficient teaching guarantees to enhance students' data mining capabilities.

Firstly, the school should update the teaching material packages. Complementary standardized resources such as project task descriptions and project instruction manuals should be added to the existing teaching task platform to help students better understand the actual content of project-based learning. For each training course, teachers should also provide students with a more detailed description of the project tasks according to the focus of the training and the evaluation criteria of the project results, so that students can train according to the task descriptions, and ensure that the project plans submitted by students can meet the requirements of the current task, so that students can exercise their practical data mining skills. Secondly, the school should establish a unified cloud experiment platform, allowing students to download effective material packages. The Python cloud platform provides a shared database for experiments, so students can avoid inconsistent data collection in version appeared in the process of operation. The platform should also upload related tools such as video course operation demonstrations to facilitate self-study by students. Thirdly, the school should utilize generative artificial intelligence to provide 24-hour online solution services for students, ensuring that students can promptly and effectively obtain solutions when encountering bottlenecks during project implementation. Teachers should also provide students with more process-oriented guidance to help them successfully complete the project. In particular, teachers should guide students to think deeply at key nodes and consider the data mining problem from the perspective of the project leader to select the appropriate algorithm. Fourthly, the school should establish a data file for each student, recording the records of each data mining training course, to help students promptly improve based on the weaknesses identified from historical data. Through targeted task completion, their data mining skills can be improved.

Through historical data, students can also independently arrange their roles in the tasks, such as data engineers and algorithm engineers, to exercise their professional skills, thereby ensuring that their abilities can be effectively enhanced. Fifth, schools should also strengthen school-enterprise cooperation, cooperate with enterprises and industry associations, provide more real projects for school students, and help students think about the specific technology application of data mining in practical projects.

5. Conclusion

This study takes the textbook *Python Data Mining Algorithms and Application Experiments and Course Training Guidance* as the core research object, and discusses the specific application of PBL teaching system in the practical teaching of big data mining. It proposes problem-oriented teaching, result-oriented teaching, scenario-based teaching, and corresponding hierarchical task teaching models. Through a systematic structure arrangement, it ensures that students can effectively participate in the design of tasks, improve their comprehensive abilities in big data mining practice, and enhance the overall quality of talent cultivation. In the future, the school should also establish the teaching feedback mechanism of multi-channel, to participate in PBL learning and conduct a comprehensive assessment of students' learning results. Based on technological development and industry needs, update the project cases of big data mining, introduce the latest algorithms and application scenarios, ensure that the teaching content remains cutting-edge and practical, and continuously innovate big data mining tools based on the latest developments in data. Establish a close partnership with enterprises, introduce practical enterprise projects, and improve the effectiveness of talent development through the integration of industry and education.

Disclosure statement

The author declares no conflict of interest.

References

[1] Chen X, Xu X, 2026, Mining Educational Big Data to Develop an Early Alert Dynamic Model of Academically At-Risk Students: A Proof of Concept. Innovations in Education and Teaching International, 63(1): 26–38.

[2] Cao Q, Jin Y, 2025, Research and Practice on Ideological and Political Teaching Reform of the "Data Mining" Course Based on the "Three-Integration and Three-Approach" Model. Journal of Research in Vocational Education, 7(9): 21–24.

[3] Cheng H, Yu B, 2026, Adaptive Teaching Mode Optimization Using Reward-Shaped Deep Reinforcement Learning and Big Data Mining. Discover Computing, 29(1): 23.