ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

# Research on Active Project-driven Teaching Practice and Teaching Reform of Data Mining Course

Qunfang Zeng[1, a], Yuan Liang[1, b, *], Zhaoyong Yi[1, c], Qiuting Du[1, d], Min Cheng[2, e]

[1]School of Management Science, Chengdu University of Technology, Chengdu 610059, China;

[2]78098 unit of Chinese people's Liberation Army, China.

[a]764321379@qq.com, [b]tacal.liang@163.com, [c]250599533@qq.com, [d]qiutingd@163.com, [e]1457274@qq.com

## Abstract

**With the development of big data and artificial intelligence technology, colleges and universities pay more and more attention to the training of talents in this field, especially offering courses in big data and artificial intelligence. As the basic knowledge of big data and artificial intelligence, the course of data mining is a subject that integrates mathematics, statistics, machine learning, information technology, etc. It is very difficult for teachers to train and students to learn. In order to improve the quality of training, the teaching method of "taking practical projects as the main line, students as the main body and educators as the leading role "is put forward. The teaching content of data mining course is integrated into specific projects, and students are guided to learn actively and autonomously, so as to cultivate a new type of big data analysis talents who conform to the era of big data and have the ability of knowledge acquisition, independent research and practical problem-solving.**

## Keywords

**Big data era, Postgraduate training, Active project-driven teaching method, Innovation.**

## 1. Introduction

The Fourth Session of the Twelfth National People's Congress pointed out that the implementation of the national big data strategic plan will accelerate the development and sharing of data resource and deepen the innovative application of big data in various industries (Liu Qun et al., 2016). Big Data technology has been rapidly applied, and industrial-scale are growing at a rate of 10-30% annually. There are nearly one-third of enterprises setting up an independent department of big data analysis for data analysis and application (Liu Bo et al., 2019). From the current data analysis, data mining technology can help companies develop business, reveal known facts, and predict unknown results. Therefore, the application of data mining technology has become a necessary method for enterprises to maintain competitiveness (Wu Huixiang, 2008). Therefore, talents in big data analysis are extremely attractive to enterprises. Facing the strong demand for big data talents by enterprises, cultivating application-oriented talents with solid data analysis capabilities and value for enterprises has become an important task for colleges and universities, and it is a huge challenge for educators to reform teaching practices (Mei Zhengyu, 2015). In order to adapt to the training of big data talents, major universities have set up courses such as "Data Warehouse and Data Mining" to allow students to systematically learn and understand knowledge in this area (Wang Jianxin, 2019).

As the "data warehouse and data mining" course itself has the characteristics of wide knowledge, great difficulty, high math requirements and strong IT requirements, educators in major colleges and universities actively combine their professional characteristics, technical

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

features from the teaching practice and reform to improve the quality of students in this regard. From the professional characteristics, data mining courses offered by economic statistics majors require students to understand what is data mining, be proficient in using typical mining tools to analyze economic indicators, and have basic ability to extract information and knowledge from the data resources and assist in decision-making (Li Zhong, 2014). A data mining course in the computer major requires students to understand the principles of different data mining algorithms and implement by programming, and learn to use data mining technology to establish mathematical models to solve practical problems. Applied statistics majors mainly require students to have the ability to use data mining software and make auxiliary decisions based on the mining results. At the same time, students need to master the principles and processing procedures of data mining algorithms, and require programming to implement algorithms. From the point of view of the quality of students, graduate students after four years of undergraduate study basically have the ability of using data thinking, and mastered a certain application development capabilities. Therefore, it is already the mainstream trend to focus on the cultivation of students' problem-solving ability and independent thinking skills. In response to this demand, changing the teaching method of the Master of Applied Statistics "Data Warehouse and Data Mining" emphasizes practical project-driven, student-centered, and educator-led changes, promotes active project-driven teaching practice, and accelerates the cultivation of students' practical ability and self-learning thinking ability (Tan Suyan, 2019).

## 2. Formulating Teaching Objectives from the Perspective of "Supply" and "Need"

The development of intelligent big data is imperative for the prosperity of the city with science and education, strengthening the city with talents. With the ultimate goal of training data mining talents with practical problem solving and exploration capabilities, it is necessary to combine the university's training program for the master of applied statistics and the hard and soft requirements for data mining talents in various fields, and targeted analysis from the two general directions of "supply" (teaching resources provided by universities) and "demand" (demands of the talent market).

### 2.1. From the Perspective of "Supply"

With the development of modern network technology, teaching is no longer limited to traditional teaching methods. More online teaching courses are respected by major universities, and online courses are highly efficient and real-time. Therefore, the rapid update of the knowledge of data mining courses needs to be closely integrated with the network, and universities can make full use of the advantages of the online real-time course to provide students with the latest knowledge. Although the advantages of online courses are obvious, they still face many challenges in the practical application of teaching in China: lack of interactive communication, lack of supervision, etc. In addition to providing online learning resources for students, universities should also provide students with a platform for mutual learning to make up for the shortage of online courses. The core goal of data mining courses is to cultivate students' ability to think hard, explore, and create. Universities can not only provide information and technical support for postgraduate courses, but also provide personalized study consultation and targeted course study guidance with regular literature reading, competitions, and academic exchanges.

### 2.2. From the Perspective of "Needs"

According to experts in the big data industry, "knowledge hidden in a large amount of data can change the world". Data mining talents act as knowledge "diggers" to replenishing humanity

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

with endless social values. To achieve this ultimate goal, all walks of life have put forward requirements for “diggers”.

Firstly, big data mining talents need to have sophisticated data thinking skills. The mature data thinking ability lies in that "miners" have a good insight into the data, and are familiar with big data technology, commonly used data mining algorithms and their application scenarios. Secondly, big data mining talents need to be proficient in operating big data commonly used software. Based on an in-depth understanding of theoretical knowledge, the final analysis results are completed through operating the tool. Therefore, statistical software such as SQL, SPSS, SAS, and a general-purpose programming language such as python are needed to be proficient in, and it also lays down a solid foundation for later organizational modeling, analysis and prediction. Thirdly, big data mining talents need continuous and rich interdisciplinary knowledge. Due to the rapid development of the big data industry, learners need to constantly improve themselves. They need to have not only data analysis, data mining, machine learning and other capabilities, but also knowledge and skills in business models, data products, etc. Fourthly, big data mining talents need to have a good sense of teamwork. The work of big data often includes two parts: data and business. A single department cannot complete all the work, so the collaboration between the data team and the business team is very important. Therefore, we should pay attention to the cultivation of communication ability, learning ability, promotion ability and execution ability in college education.

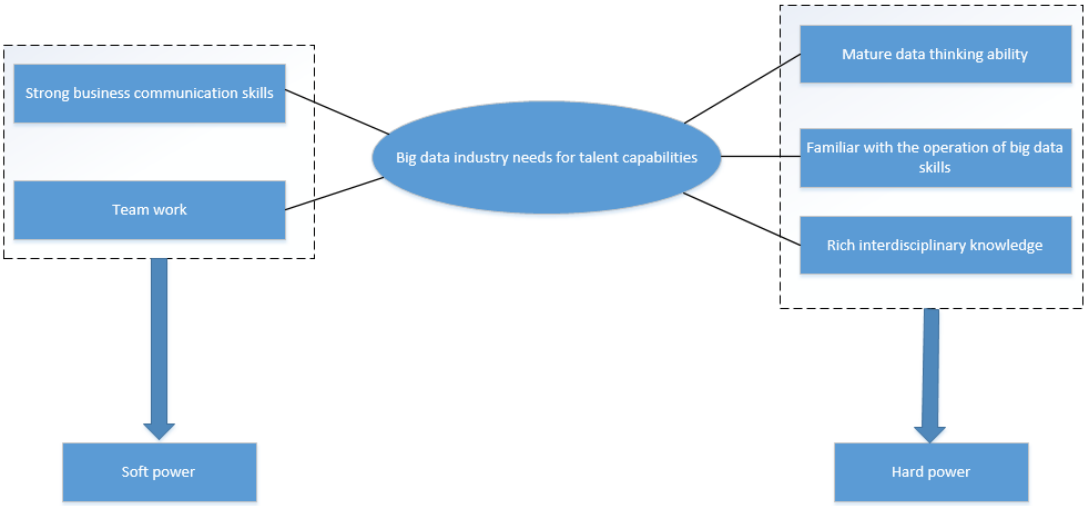


**Fig 1.** Quality requirements for big data talent training

## 3. Designing Data Mining Course under Project-Driven Teaching

Data mining courses focus on the combination of theory and practice, and educators have conducted relevant research on the content of data mining courses from different angles. Some scholars have studied the experimental content of the data mining course based on the different programming languages used in the course (Huang Lan&Zhou Juan, 2017; YUE Qiang et al., 2016). Some scholars have discussed the problems encountered in the teaching process of the data mining course (Luo Shiguang et al., 2013; Wang Ruixiang, 2017). There are also some scholars who have studies how to build a data mining platform (Xie Yaobing et al., 2014; Guo Xin et al., 2014). But these studies only give students a shallow understanding of the process of data mining and the principles of algorithms, but they do not actually allow students to appreciate the application of data mining courses in a certain industry.

The Ministry of Education-Beijing Nazhi Industry-Academic Cooperation Collaborative Education Project is funded by Beijing Nazhi Campus Technology Development Co., Ltd. It aims at assisting partner universities in curriculum construction through the establishment of a

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

curriculum funding project to improve teaching quality and promote compliance with industry needs Talent training. By allowing students to research and develop projects based on big data technology to train student's innovation and entrepreneurship ability and practical ability, improve their comprehensive literacy and require students to complete innovation and entrepreneurship training programs on time and with quality. After the one-year project, students are required to submit the required development results. This project, led and planned by the Education Committee's cloud computing expert group, covers demonstration courses and teaching materials for cloud computing, big data and cloud computing security, fully integrates the advanced concept of the industry's first-line practice, and promotes the reform of college courses. Referring to the talent training plan of the Ministry of Education-Beijing Nazhi Industry-University Cooperation Collaborative Education Project, exploring the teaching content driven by the active project of the data mining course, and improving the teaching content of the combination of algorithm principles and practical applications.

The first level is the data cleaning process. Data cleaning is mainly to delete irrelevant data in the original data set, duplicate data, smooth noise data, and filter out data irrelevant to the mining theme, deal with missing values, outliers, etc. These steps require students have a more intuitive understanding of the common methods of data types and missing values filling. After learning the theoretical knowledge of data processing and writing code, students are proficient in applying relevant methods to lay the foundation for the actual operation of subsequent projects. Data cleaning is an indispensable step in the data analysis process. The quality after data cleaning directly affects the quality of our subsequent work, so students need to pay special attention to data cleaning. Therefore, after completing the introduction of the method, educators should find relevant cases for students to carry out practical operations, so that students can achieve the purpose of understanding data cleaning methods and principles.

The second level is algorithm principle verification. Students use Python or R programming language to program common mining algorithms in data mining courses to realize, analyze and verify their principles. There are many data mining algorithms. Educators divide various algorithms according to different classification standards, and introduce algorithm principles according to the order of the classification framework. Commonly used unsupervised learning algorithms: K-means algorithm, auto-encoder, principal component analysis, commonly used supervised learning algorithms: K nearest neighbor, decision tree, support vector machine, naive Bayes, etc. Among them, unsupervised learning is mainly to solve the problem of clustering, and supervised learning is mainly to solve two problems: regression, classification, corresponding to quantitative output and qualitative output, respectively. Any teaching content must be based on a certain knowledge classification theory. In the face of different types of knowledge, the form and method of instructional design should also be different. If all adopt the same method for teaching design and adopt the same model for teaching, one can imagine that the teaching effect will not be good. Therefore, scientifically and rationally classifying knowledge is helpful for rational instructional design and affects the quality of instruction to a large extent. Therefore, putting the same category of algorithms into the same case to verify the principles will help students have a deeper understanding of the similarities and differences between the same category of algorithms.

The third level is the application of data mining scenarios. This stage runs through the entire data mining course learning process, which helps students to use the knowledge of each learning stage proficiently and improve their practical ability and exploration and thinking ability. The cultivation of professional postgraduates should be based on practical projects, and divide students into groups that undertake different types of tasks, including information collection groups, data cleaning groups, data fusion groups, data mining groups, and data visualization groups. The information collection group mainly collects data through web crawlers. It can also use other methods to collect data according to business needs. The data

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

cleaning group mainly finds out invalid "dirty data" and removes or replaces it. The task of the data cleaning group is very heavy, and the work cycle is usually very long. The data fusion group mainly is to classify and organize the crawled data information. If there is a superior-subordinate relationship, it is listed according to the relationship between the subclass and the parent class. The data mining group is to get the processed data and use the data mining algorithm to study the causal relationship between the previously set impact factors. The main classification algorithms are decision tree, Bayesian classification, rule-based classification, neural Networks, support vector machines, K-nearest neighbor classification and case-based reasoning in lazy learning algorithms. As the name implies, the data visualization group visually displays the results of the data mining group and presents the unknown information found through data analysis. In this way, the practically valuable information can be captured intuitively. Data mining projects are completed by teamwork, and the quality depends on the depth and breadth of students' data mining activities. The entire team can have different organizational forms and different affiliated relationships, but it is necessary to communicate with each other to achieve the purpose of learning and solving practical problems together.

## 4. Positioning Roles under Project-Driven Teaching

(1) Educators guide learning

Educators are not only the imparters of knowledge but also the guides of teaching. They should also guide the development of the entire teaching, and guide students to find more problems, think more, and find ways to analyze and solve problems (Hu Wei & Yan Ming, 2005). The introduction of active project-driven teaching methods will change the traditional teaching methods, so the knowledge and abilities that educators need to possess are also very different from the traditional teaching methods. Therefore, under the adaptive project-driven teaching method, educators need to do the following three aspects.

Firstly, educators need to optimize and implement teaching projects. In order to promote the efficient operation of project-driven teaching methods, educators need to optimize the design of teaching projects and ensure the smooth implementation of project activities. In actual operation, educators should consider both course content and students' learning abilities, and select suitable teaching content as much as possible to promote the effectiveness and maximization of students' knowledge of the curriculum. In addition, in order to ensure the stable operation of the project and mobilize the enthusiasm of students, educators should consider the synchronization of teaching projects and teaching processes, and ensure that the design projects are close to the lives of students.

Secondly, educators need to condensate the course content. Due to the wide range of data mining courses and complex principles established by the project-driven teaching method, the implementation of condensed course content often has higher requirements for the professional level and ability of educators. Before teaching, teachers should analyze the general framework of the teaching content and ingeniously condense the course content according to scientific standards so that students can master the course content efficiently.

Thirdly, educators need to guide students to change their learning methods. Under the project-driven teaching method, students need to simulate the actual project application scenario, and each team collaborates and communicates with each other to achieve the purpose of learning. Therefore, the course teaching activities should simplify the teaching content of educators in the process of carrying out, thereby promoting students' independent inquiry and improvement of learning ability.

(2) Active learning of students

The recipients of knowledge are students, but the effect of students' knowledge acquisition is different due to different learning methods. American scholar and famous learning expert

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

Edgar Dale first discovered and proposed the "learning pyramid" theory in 1946. The theory proposes that the learning effect of traditional passive learning or individual learning is less than 30%, while the learning effect of team learning, active learning and participatory learning is more than 50% (Gao Feng, 2019). Team learning, active learning and participatory learning in project-driven teaching include reading literature, participating in discussions, speaking, hands-on practice, making reports, and imparting to others. The whole process of the project-driven teaching method guides students to adjust their learning methods according to the "learning pyramid" theory, and its learning effect is much higher than the traditional passive learning method. The retention rate of "learning pyramid" theoretical learning content is shown in Fig.-2.

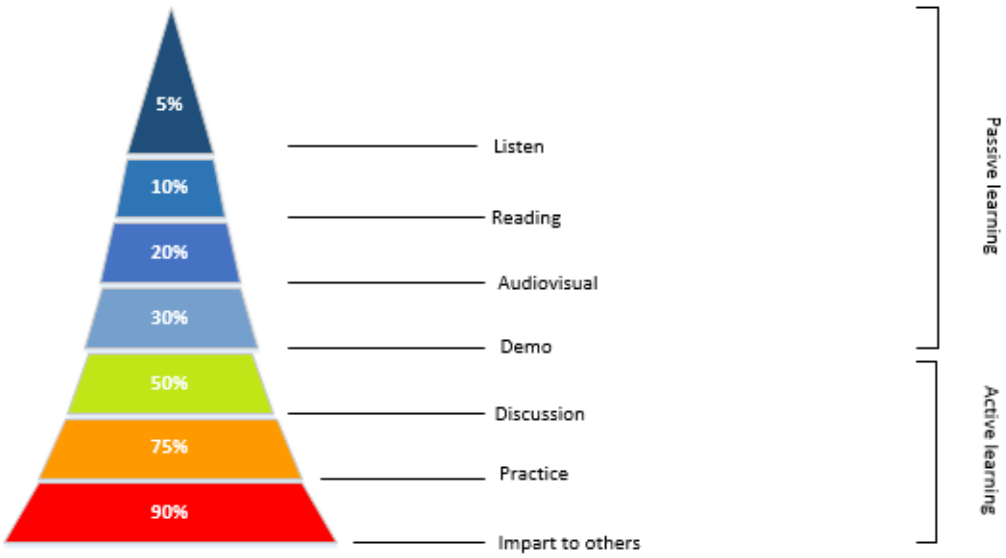


**Fig 2.** Learning pyramid

## 5. Conclusion

In order to effectively stimulate postgraduates' interest in learning and enhance their self-directed learning ability, taking the data mining course as an example, the teaching objectives are rationally formulated by combining the "supply" of universities and the "demand" of the talent market, and the content characteristics of data mining courses, innovated the teaching content driven by the active project. This innovative project-driven teaching method fundamentally puts forward higher requirements for the knowledge and ability that educators need to possess, and also guides students to change from "passive learning" to "active learning". The teaching method is mainly innovative, practical, and effective, following the principles of teacher guidance and student active learning, thereby effectively improving teaching efficiency and quality, and training students' independent research ability and solving practical problems, and training students to develop the habit of active learning and lifelong learning, which makes the graduate students trained in colleges and universities have a sustainable learning motivation. Under the pressure of fierce employment competition, colleges and universities can better reserve higher quality college graduates for the talent market.

## Acknowledgments

This research was financially supported by Chengdu University of technology 2018-2020 higher education talent training quality and teaching reform project.

ISSN: 2637-6067 DOI: 10.6918/IJOSSER.202009_3(9).0041

"Research and Practice on teaching mode reform from teaching to introducing learning under the background of Internet plus" (Grant No.JG183090) and the Project of Graduate Education Reform Program at Chengdu University of Technology "Research on active project-driven teaching practice and reform in Data warehouse and data mining" (Grant No. 10800-00009825).

## References

[1] Liu Qun., Wang Guoyin., Yu Hong., & Xia Ying. (2016). Discussion on classroom teaching practice combined with online course education——Taking the basic course of data mining as an example [J]. Computer Education, (10): 52-54 + 59 .

[2] Liu Bo., Cai Yansi., & Zhong Shaodan. (2019). Exploration of practical teaching of data mining courses under the background of big data [J]. Journal of Higher Education, (18): 124-125 + 128.

[3] Wu Huixiang. (2008). The application of data mining in the financial risk early warning system [J]. Finance and Accounting Newsletter (Comprehensive Edition), 2008 (02): 91.

[4] Mei Zhengyu. (2015). Security Big Data Analysis: Building an Enterprise Security Ecology [J]. Software and Integrated Circuits, (09): 28-31.

[5] Wang Jianxin. (2019). Exploring the training methods of data thinking in data mining courses [J]. Computer Education, (08): 117-120.

[6] Li Zhong., & Li Shanshan. (2014). Colleges of applied IT professional data mining course construction [J] computer. Generation, (11): 65-66 +69 .

[7] Tan Suyan.(2017). The construction of the "three main" flipped classroom teaching model supported by MOOC [J]. Contemporary Education Theory and Practice, 9 (11): 80-83.

[8] Huang Lan., & Zhou Juan. (2017) Experimental design of data mining course for applied undergraduates based on Weka [J]. Computer Knowledge and Technology, 13( 21 ): 7-9.

[9] YUE Qiang., Huzhong Yu., & Wen Jin.(2016). Experimental design of data mining course based on R language [ J ]. Microcomputer Application, 32 (5): 31-37.

[10] Luo Shiguang., Luo Changri., & Zhang Xinhua.(2013). Experimental Teaching of Financial Data Mining Based on Comprehensive Design Experimental Project [J]. Teacher Education Forum, 2013 , 26 (3): 83-87.

[11] Wang Ruixiang.(2017). The application of anchored teaching in the experimental teaching of data mining in TCM schools [J]. Western Medicine, 30 (4): 54-55.

[12] Xie Yaobing., Xiang Yang., & Huang Zhenhua. (2014). Construction of experimental platform of data warehouse and data mining course based on Android mobile terminal [J] . Education Teaching Forum, (6): 248-249.

[13] Guo Xin., &Chen Qian. (2014) Research on the construction of data warehouse and data mining teaching experiment platform [J] . Education Teaching Forum, (9): 253-255.

[14] Hu Wei., & Yan Ming. (2005) The application of "independent" classroom teaching mode in physics teaching [J]. Scientific Consultation (Education and Research), (11): 54.

[15] Gao Feng. (2019) Research on the improvement of middle school physics teaching based on the "learning pyramid" [J]. Fujian Basic Education Research, (10): 113-115.