Universe Scientific Publishing

DOI:10.18686/ahe.v7i12.8321

Construction of Course of Introduction to Data Science

Changqing Yu[1*Corresponding author], Ji Chen[2], Wei Wu[1], Yarong Chen[1], Peng Chu[1], Kai Yang[1]

1 Xijing University,Xi'an,Shaanxi,710123 ,China.

2 Chongqing Hanhai Ruizhi Big Data Technology Co.,Ltd.Chongqing,400000 ,China.

Abstract: With the advent of the era of big data,data has become an important asset in various industries,and the demand for data talents in society is also increasing.Universities have become the main battlefield for talent cultivation.Introduction to Data Science is an important basic course involving multiple and highly interdisciplinary emerging disciplines.Based on actual teaching feedback,this article discusses the teaching of the Introduction to Data Science course from aspects such as teaching content settings,teaching methods,and provides exploratory suggestions for cultivating data talents.

Keywords: Introduction to Data Science;Course Construction;Course Design

Fund Project:

Xijing University's third batch of high-quality undergraduate courses and key course construction projects(XJZDKC22009 XJZDKC22010 XJJPKC22008);

The Ministry of Education's industry-university collaborative education project(220603414290713);

The 2021 Undergraduate Teaching Reform Research Project of Xijing University(JGGH2103);

Xijing University Teacher Education Reform and Teacher Development Research Project(JFGG202203);

Special project for undergraduate talent cultivation program research of Xijing University(exploration and practice of cross integration and support construction of big data majors of Xijing University);Shaanxi Undergraduate and Higher Continuing Education Teaching Reform Research Project(21BG051 21BY180);

2021 Annual Project of Shaanxi Provincial Education and Science"14th Five Year Plan"(SGH21Y0283)

Introduction

With the advent of the era of big data,data has become an important asset for various industries,and the demand for data talents in society is also increasing.Universities have become the main battlefield for talent cultivation.Introduction to Data Science is an introductory course that involves a highly interdisciplinary and emerging discipline.It is an introductory course designed to provide students with an intuitive understanding of the development history,knowledge structure,training objectives and requirements of data science and big data technology,computer science and technology majors,as well as basic knowledge,typical technologies,and specific applications related to data science and big data technology,to help undergraduate students understand the basic overview of their majors,establish correct professional ideas and master university learning methods.

1. Content of Courses

1.1 Course system architecture

The teaching content of the course revolves around the overall structure of the professional course system.For the professional course system,the professional introduction course needs to connect the main knowledge points of the major,not only extracting key knowledge from each major course for comprehensive explanation,but also taking into account the relationship between various knowledge,to enable students to fully understand the background and development history of data science and big data technology,professional characteristics and comprehensive requirements,as well as relevant complete knowledge and skill systems. It explains the correlation between majors closely related to data science and big data technology,such as computer science and

technology,statistics,etc.,and expounds the talent demand,job positions,and employment situation of data science and big data technology majors.

1.2 Course theoretical content

According to the nature and status of the curriculum in the talent cultivation plan,the curriculum teaching content is arranged according to the curriculum syllabus.Firstly,it introduces the development history,professional knowledge requirements,and basic skills requirements of data science and big data technology.Secondly,it introduces the basic concepts of data science,the basic methods of data mining,and the main techniques of big data analysis.Next,it introduces the basic process of data science,the ecological environment of big data technology,and the key technologies of big data platforms.Finally,based on actual projects,examples such as medical big data and intelligent urban transportation big data are introduced.

1.3 Professional skill system

The professional skills system comprehensively considers the actual needs of society for computer and big data professionals. Based on the accumulated professional construction experience,students majoring in computer science and technology,data science,and big data technology need to learn various technologies and skills throughout the entire process from data acquisition to data analysis and application.For the introduction course,students are not required to master all the technologies,but they need to be familiar with the technical routes of data science and big data,be able to select corresponding technologies for key links in the route and consult materials for further learning,and have the ability to solve practical problems.

2. Teaching Model

Based on the MOOC of Introduction to Data Science online by the school,a blending learning model of SPOC(Small Private Online Course)is developed combining classroom teaching and online teaching.First,online teaching resources need to be arranged for students in the school.Then,in offline classroom teaching,students'questions should be answered to understand what knowledge they have absorbed and what has not yet been absorbed.Teachers can work with students on homework or other tasks in class.Finally,through the systematic summary after class,further consolidation and improvement are made.Teachers can freely set and control the progress,rhythm,and scoring system of courses based on students'needs.Students must ensure learning time and intensity,participate in online discussions,complete prescribed assignments,and exams.Those who pass will receive a course completion certificate or credit.

3. Teaching Methods and Means

Taking students as the center,we should change from"teaching"to"learning",from"listening"to"asking",from"learning"to"resear ch",so that not only"candidates"can become students,but also teachers can become"mentors",and"textbooks"can become"reference books".

3.1 Teaching Learning Methods

Students'interest in learning should be stimulated and self-confidence in learning should be established.Using the"virtual reality and metauniverse"approach to project driven tours of data science,students are guided to summarize the rules,processes,and skills that should be mastered in the course of data science.Combined with big data and artificial intelligence competition questions,students are given the key to opening the door to data science,allowing them to enjoy themselves in the palace of data science.

3.2 SpocTeaching

Modern information technology can be used to introduce SPOC teaching.Before class,teachers are learners and integrators of curriculum resources,integrating online and physical resources according to students'needs.In the classroom,teachers are mentors and facilitators,organizing student discussions,providing personalized guidance at any time,and working together to solve problems encountered.After class,customizing courses for students and providing them with differentiated and stronger professional support can enhance their complete experience of the course.

3.3 Combining discussion and communication with flipped classroom

Each chapter requires students to read classic English materials,such as Google's three classic papers,introducing GFS,MapReduce,and BigTable,the pioneers of cloud computing technology and big data technology,and producing PPT(power point)that are flipped and discussed in class.

3.4 Research

For each knowledge module,students should be assigned after class research questions.In combination with the current real-time hot issues,data science methods should be used to conduct research,and form research literature to upload to the school cloud.At the

same time,this can serve the purpose of ideological and political courses.

3.5 Immersive experiential programming

Through the Aishuke platform,using“drag and drop”programming,basic data science skills such as data preprocessing,data analysis,data mining,and data visualization can be experienced,thereby enjoying the great charm of data science with”what you see is what you get”.

3.6 Project guide

Using cases such as the Titanic case,the 2012 Obama presidential election,and the production of hot word maps for the 2017 Central Spring Festival Gala,students will gradually learn about the application of data science skills through Python and Spark programming.

4. Practical Teaching Resources and Teaching Conditions

Special experimental guidance books have been prepared for Introduction to Data Science.A total of 8 experiments have been prepared,including 4 required lectures and 4 optional lectures,which not only meet the learning needs of most students,but also enhance the learning needs of students with strong learning initiative through selective experiments.All experiments have been conducted in the engineering boat,basically meeting the experimental teaching requirements.We should make full use of network experimental resources and actively explore the EduCoder Experimental Platform,Alibaba Tianchi Experimental Platform,Aishuke Experimental Platform,PaddlePaddle Experimental Platform,etc.We also need to improve the experimental environment and resources of the Internet of Things and big data research platform.Alibaba AI Data Science Training Camp-Xijing University Station provides jumping experiment content for top students.These students can log in to the experimental platform for improvement training anytime and anywhere.

5. Course Assessment

The assessment method for students in this course consists of their usual scores(online+offline)and their final exam scores,among them,chapter learning score(watching video)5%+homework score(thinking questions,discussions,research and analysis,quizzes,moral education and politics,and regular bonus points)25%+classroom score 40%+final test score 40%.At the same time,through the analysis and mining of learning process data,and taking into account the comprehensive evaluation of students’performance in online communication,participation,and effort level,the overall evaluation score is fine-tuned to fully reflect the efforts and gains of students.

6. Teaching Effects

Was approved by the Ministry of Education for the construction of the“Introduction to Data Science”curriculum system for collaborative education between industry and education;established a data science club with Alibaba to allow students to participate in the Dragon Ball Training Camp held by Alibaba’s Tianchi Big Data Research Platform,two winter vacation charging plans,and a Tianchi school start gas kit;organized the 4th National Communication Intelligence Cup IT Skills Competition for College Students,with 1 first prize,5 second prizes,and 17 third prizes;participated in the fourth“Teddy Cup”data analysis skills competition,with 1 first prize,7 second prizes,and 2 third prizes;won the the Belt and Road and BRICS Skills Development and Technology Innovation Competition Machine Learning and Big Data Application;won 1 national third prize and 1 national excellence award;participated in the Shaanxi Telecom and Internet Industry Vocational Skills Competition and the Fourth National University Student Big Data Skills Competition,and won one provincial second prize,one provincial third prize and two outstanding instructors.

References:

[1]Chao Lemen.Open Source Curriculum and Introduction to Data Science[J].Computer Science,2020,47(12):114-118.

[2]Zhou Jianying,Li Guangming,Wang Shuangcheng.”Introduction to Data Science”Course Teaching Discussion[J].Education and Teaching Forum,2020(24):230-231.