DOI: 10.23977/curtm.2025.080414
ISSN 2616-2261 Vol. 8 Num. 4

Instructional Design and Exploration of Big Data Fundamentals Course for Non-Computer Science Disciplines

Li Wang

School of Economics, Nanjing University of Finance and Economics, Nanjing, Jiangsu, 210023, China
3790792758@qq.com

Keywords: Interdisciplinary education, big data foundation courses, non-computer science disciplines, data technology application ability, Python programming practice

Abstract: Amid the rapid advancement of the digital economy, big data technology has emerged as a core driver of industrial transformation. However, big data courses for non-computer science disciplines face persistent challenges, including students' weak technical foundations, disconnects between curricula and discipline-specific demands, and insufficient practical training. This study proposes a systematic pedagogical reform framework grounded in interdisciplinary education theory, integrating constructivism and CDIO engineering education models. The framework establishes a scenario-based teaching system with three pillars: (1) Modular content architecture ("foundation-core-extension") featuring cross-disciplinary integration of computer science, statistics, and domain-specific knowledge (e.g., medical imaging analysis); (2) Hierarchical teaching strategies with dynamic remediation mechanisms (e.g., Blockly-based visual programming for humanities students, API development for engineering cohorts); (3) A three-dimensional evaluation model (process-result-development) incorporating discipline-tailored assessments (financial analysis for business majors, sentiment analysis for humanities, environmental modeling for STEM fields). Practical implementation employs enterprise collaboration platforms for authentic case simulations (e.g., e-commerce user behavior analysis), cloud-native environments (Kaggle/AliCloud), and lightweight tools to strengthen full-process data competencies. The reform embeds data ethics education and ideological elements through privacy protection modules and scenario-based decision-making exercises. By aligning pedagogical design with industry needs and leveraging cloud-based resource integration, this framework provides a replicable model for enhancing non-computer science students' analytical capabilities and interdisciplinary literacy in big data education. Future research will explore AI-driven adaptive learning systems and longitudinal graduate competency tracking to optimize curricular efficacy.

1. Introduction

With the rapid development of new generation information technologies (big data, artificial intelligence, cloud computing, blockchain), the digital economy has become a new engine driving high-quality economic development. Big data applications have penetrated into fields such as finance, healthcare, and education, and many companies have listed data analysis skills as a core job requirement. In this context, the interdisciplinary nature of big data foundational courses, as the core carrier for cultivating digital literacy, is increasingly prominent.

Non-computer science students generally have weak programming foundations and insufficient knowledge of mathematics and statistics, making it difficult for them to quickly master data analysis skills. Although big data related courses have been widely offered in domestic and foreign universities, there are still the following problems in teaching for this group: 1) unclear course positioning: most courses directly transplant teaching content from computer majors, lacking targeted design for the characteristics of non-computer science students; 2) The disconnect between theory and practice: The teaching content focuses more on theoretical explanations, with insufficient practical activities, making it difficult for students to transform knowledge into practical application abilities; 3) Lack of disciplinary characteristics: The course content has not been deeply integrated with various professional fields, making it difficult to meet the actual needs of students from different majors; 4) The evaluation system is single: overly relying on final written exams and lacking effective assessment of practical abilities.

Big Data: A Revolution That Will Transform How We Live, Work, and Think The "full sample analysis" thinking paradigm[1] proposed in the book laid the theoretical foundation for this course. As a big data foundation course aimed at non-computer science students, this course deeply integrates interdisciplinary knowledge systems such as mathematics, statistics, computer science, and machine learning, systematically teaching key technologies such as data acquisition, distributed storage, and parallel processing. The course incorporates data analysis knowledge and skills into the training program requirements, helping students build a complete transformation chain from raw data processing to decision support, aiming to enhance the data analysis ability and comprehensive quality of non-computer science students, and equip them with the core competitiveness required in the era of big data.

2. Overall Instructional Design

2.1 Curriculum Design Concept and Framework System

2.1.1 Interdisciplinary Positioning and Ability Development Objectives

Driven by digital transformation, this study aims to construct a big data basic curriculum system for non-computer science disciplines (including humanities, engineering, management, and other disciplines). The curriculum architecture is based on constructivist theory (emphasizing collaborative knowledge construction and socio-cultural reflection) [2], combined with the practical guidance framework of CDIO engineering education model [3], focusing on cultivating new talents with triple composite abilities: data technology application ability, interdisciplinary integration ability, and data ethics decision-making ability.

The core positioning of the course reflects three characteristics:

1) Cross disciplinary integration: Breaking through the traditional boundaries of a single discipline, establishing a diverse knowledge fusion of computer science (HDFS/MapReduce principles), statistics (data modeling methods), and professional domain knowledge (such as medical imaging analysis, social science surveys);

2) Scenario based application: Emphasize the ability to solve problems in real scenarios, design a full chain practical project covering data collection, cleaning, and analysis;

3) Universal adaptation: Adapt to the needs of different disciplinary backgrounds in humanities,

sciences, and engineering, ensuring compatibility between course content and knowledge structures in different disciplines.

Based on Bloom's educational goal classification theory [4] and ACM's data science curriculum standards [5], a four-dimensional ability evaluation system is constructed, and empirical teaching strategies are integrated to achieve multidimensional ability collaborative development. The cognitive dimension focuses on the construction of a knowledge system of technical principles, requiring students to master the core features of big data, distributed computing principles (including MapReduce parallel framework and HDFS storage architecture), etc., and deepen their understanding of distributed storage and computing paradigms through experimental platforms (such as HDFS architecture visualization interaction system). The tool dimension focuses on strengthening the application capability of the entire process data analysis toolchain, with capability indicators covering key links such as data organization (Excel), visualization analysis (Tableau), programming modeling (Python), and cloud collaboration (Kaggle). In teaching implementation, an end-to-end practical project-based model is adopted to guide students to gradually advance from data cleaning and feature engineering to cloud collaboration modeling, with a focus on cultivating the organic integration ability of the toolchain. The transfer dimension emphasizes the transformation of interdisciplinary problem-solving abilities, requiring students to achieve a closed-loop mapping of "problem → data → solution", and design interdisciplinary scenarios through project driven teaching to promote the transfer of data thinking to professional practice; At the emotional level, the course focuses on shaping the decision-making and judgment of data ethics and privacy protection. Through case studies and scenario simulation teaching, students are guided to dialectically evaluate the balance between data utility and privacy risks.

2.1.2 Hierarchical Admission Mechanism and Teaching Adaptation

The course sets Python programming basics as a mandatory prerequisite, with specific requirements including: the ability to build a development environment (Anaconda configuration success rate ≥ 95%); Mastery of core grammar (accuracy rate of loop/function structure ≥ 80%); Proficiency in calling commonly used libraries (pandas basic operation completion time ≤ 15 minutes/task).

Design a dynamic compensation teaching system based on Vygotsky's zone of proximal development theory [2] to address differences in subject cognition, ensuring that students from all majors can effectively enter the learning state. Humanities compensation layer: using Blockly graphical programming to reduce logical barriers, such as designing data cleaning processes through drag and drop modules; Science reinforcement layer: Strengthen the understanding of data structures, such as using Excel matrix operations to simulate Hadoop data processing logic; Engineering advancement: Emphasis on API interface calling, such as conducting Postman development training.

2.2 Course Content Architecture

2.2.1 Modular Design

The course adopts a three-tier architecture of "foundation core expansion", with each module consisting of three components: theoretical explanation, case analysis, and practical projects. The basic module (6 class hours) includes an overview of big data technology (2 class hours) and data organization methods (4 class hours); The core module (30 class hours) includes data analysis and modeling (12 class hours), visualization technology (6 class hours), and comprehensive project practice (12 class hours); The expansion module (12 class hours) includes machine learning applications, ethical and regulatory discussions, and competition guidance.

2.2.2 Key Technology Teaching Content

(1) Overview module of big data technology

As the basic introductory module of this course, the overview of big data technology aims to help non-computer science students establish a comprehensive understanding of big data technology, cultivate students' mastery of cloud computing, big data, data science, and related theoretical and technical concepts. The content covers data lifecycle management (collection, storage, analysis), distributed computing frameworks (such as Hadoop, Spark), and typical technical frameworks (data warehouses, visualization platforms). Starting from daily life scenarios (such as personalized recommendations for online shopping, real-time traffic scheduling, etc.), avoiding the accumulation of professional terminology, explain the 5V characteristics of big data in a simple and understandable way: large data volume, fast speed, multiple types (including text, images, videos, etc.), low value density (requiring the extraction of useful information from massive data), and high authenticity requirements [6]. This module will showcase the evolution of technology based on the time dimension: from early Excel data processing to supermarket shopping analysis systems, and now to big data tracking technology used in epidemic prevention and control. The explanation of the technical framework will avoid the accumulation of professional terminology and focus on the three-layer application architecture: the bottom layer of data collection (such as sensors and APP logs), the middle layer of data processing (such as Hadoop and Spark), and the top layer of intelligent applications (such as personalized recommendations and risk warnings). Through practical cases and visual demonstrations, we aim to help students from non-computer science disciplines understand the basic logic and application value of big data technology, cultivate data thinking, and lay a foundation for future learning of big data applications in various industries.

(2) Data organization method module

The teaching objective of this module is to cultivate the ability of non-computer science students to complete the entire process from multi-source data collection to ready data analysis. Ultimately, they will be able to: ① identify four types of data from web pages, apps, etc. ② integrate data from at least two databases using Python tools ③ complete standardized preprocessing of raw data, including cleaning and conversion.

The key knowledge points of data collection technology include: comparison of data types (structured, semi-structured, unstructured, real-time streaming), multiple terminal data sources (web pages, apps, IoT devices), and selection of collection tools (crawlers, APIs, sensor protocols). In the practical stage, use cloud platforms (such as Kaggle) or Python crawler frameworks to complete data collection practice.

The collection of big data is usually received from multiple terminals such as web pages, mobile apps, and IoT devices, and stored in multiple databases. It is necessary to integrate all data into a distributed database or cluster. The knowledge points of multi-source data integration include differences in database types (relational databases and NoSQL databases) and the principle of distributed storage (HDFS/cloud storage). In the practical stage, students are required to store data from different databases (such as MySQL and MongoDB) uniformly in a distributed cluster, with the goal of cultivating their ability to integrate multi-source data and operate data in databases.

The collected multi-source data cannot be directly analyzed and mined, usually requiring a data preprocessing process. The key knowledge points of data preprocessing include data cleaning (such as handling missing values and outliers), data transformation (such as standardization and discretization), etc. Through practical exercises with Python libraries, students will be able to complete the complete preprocessing process of raw data, with a focus on cultivating their ability to diagnose problems when dealing with actual business data.

Each step is equipped with real dataset exercises, such as COVID-19 pandemic data , A-share trading data [7].

(3) Data Analysis and Modeling Module

This module teaches data analysis methods and model construction techniques, aiming to cultivate students' ability to apply advanced data science technology to solve complex business problems. Data analysis includes descriptive analysis, diagnostic analysis, predictive analysis, prescription analysis, unstructured data processing, multimodal analysis methods such as images and speech. The content of data mining tools includes graphical interface tools (Weka, SPSS Modeler, etc.) and programming tools Python; The content of data mining includes classification, clustering, association rules, and time series prediction. It is also necessary to understand the commonly used algorithms and models for advanced applications of data analysis such as machine learning and artificial intelligence, such as decision trees, support vector machines, neural networks, etc. By using lightweight tools and avoiding complex programming and system architecture knowledge, with knowledge learning and data modeling as the core, non-computer science students can quickly master the core skills of big data analysis, thereby creating and improving production value.

(4) Application of Data Visualization in Big Data Environment

The content of this module includes the basic knowledge, methods, and tools of big data visualization, cultivating students' ability to display data through charts and other forms. Introduce low code visualization tools such as Tableau and Looker Studio (formerly Google Data Studio). Introduce Python based visualization tools, such as matplotlib and Seaborn library, which support conventional charts such as line charts and heat maps; Plotly combined with Dash framework can create interactive charts (such as 3D scatter plots, dynamic dashboards), suitable for advanced teaching scenarios; PyEcharts provides rich interactive chart templates with abundant example code and chart cases, making it convenient for students to quickly learn and reference, reducing the exploration time in the development process.

(5) Ideological and political elements

In order to implement the fundamental task of cultivating morality and talents, the curriculum will integrate ideological and political education throughout the entire teaching process. In the teaching process, by teaching about China's contributions in the history of technology (such as Alibaba Cloud technology), interpreting industry norms and standards, and selecting typical technological events such as data privacy breaches and other real-life scenarios, a teaching system that integrates knowledge imparting and value guidance is constructed to guide students to establish correct values and professional ethics; Through independent exploration and group collaboration in practical teaching, cultivate students' craftsmanship spirit and teamwork awareness; Through the design of multi domain big data processing solutions, we aim to enhance students' technical abilities while also helping them develop systematic thinking and establish a correct view of the big picture and responsibility; Teachers need to strengthen their own self-cultivation, strictly regulate themselves in academic norms, engineering ethics, and other aspects. While enhancing their professional abilities, they should also improve their ability to integrate ideology and politics, and play an active role in teaching by example in ideological and political education. Through subtle influence, they can influence students with the correct role model and personality charm, providing an effective path for cultivating digital technology talents with both professional technical abilities and social responsibility awareness.

3. Teaching Implementation Strategies

3.1. Strengthen the Teaching Staff

The rapid iteration and interdisciplinary integration of big data technology have put forward higher requirements for the professional competence and practical ability of the teaching staff. To cope with the dual challenges of accelerated technological changes and the widening gap between industry and academia, it is necessary to enhance teachers' comprehensive abilities from the dimensions of knowledge and tool updates, industry education integration, etc. Specific measures are as follows: teachers need to actively follow the development trend of big data, not only master big data related knowledge and technology, but also be good at applying big data tools and methods to solve problems; By conducting teaching evaluations, certification training, and teaching competitions, one can enhance their professional abilities. At the same time, practical skills can also be improved through enterprise internships, training, and further education; In addition, schools can also introduce enterprise engineers and teachers through school enterprise cooperation to form a hybrid teaching team, forming a "theory+practice" joint teaching and research group, and jointly developing integrated courses between industry and education.

3.2. Implementation of Intelligent Blended Learning

Relying on the national smart education platform, rich online teaching resources such as MOOCs, and diverse learning materials provided by learning communities, students can explore knowledge and hone skills on their own; Utilize teaching platforms such as Chaoxing Learning Platform and Rain Classroom to achieve full process digital management, enabling interaction between teachers and students in teaching tasks, such as online publication of learning materials, assignment of homework, and collection of homework; Utilize social media platforms such as WeChat and QQ to establish instant communication channels, increase the frequency of teacher-student interaction, and bring teachers and students closer together. By integrating online and offline teaching modes, data analysis of students' learning behavior is conducted, teaching strategies are dynamically adjusted, and precise teaching and intelligent management are achieved.

3.3. Strengthen Experimental and Practical Training Activities

Big data technology is widely applied in various fields such as social media, biomedicine, travel and tourism, financial analysis, and mechanical manufacturing. To enhance the practical ability of non-computer science students, cultivate their ability to flexibly apply big data technology frameworks, guide them to analyze corresponding data, interpret results, predict the future, and use the analysis results to form conclusions, forming a relatively complete understanding of data science and big data technology system, and effectively cultivating students' full chain skills in data collection, cleaning and processing, data analysis, data mining, visualization, data reporting, etc.

Integrating big data technology examples from various fields into the teaching process, setting basic and comprehensive cases based on the complexity of the cases. Basic cases are generally lightweight practices that individuals can complete in a single field, focusing on the cultivation of basic abilities throughout the entire data process; Comprehensive cases are generally cross disciplinary practices completed by groups, strengthening technical integration and systematic thinking.

By collaborating with leading enterprises to build training platforms in specific fields, introducing real business scenarios of enterprises (such as e-commerce user behavior analysis), adding data privacy protection modules, integrating ideological and political education, constructing training platforms, simulating real practical activities such as data collection, cleaning and processing, analysis, etc., and presenting the results in a visual way, we jointly cultivate high-quality applied talents that meet the needs of industry development.

3.4. Strengthen Assessment and Evaluation

To effectively evaluate students' mastery of knowledge, learning outcomes, and problem-solving abilities, this course adopts a three-dimensional evaluation model to achieve dynamic monitoring and accurate feedback. Through clear evaluation criteria, teachers can promptly identify students' learning shortcomings, adjust teaching strategies in a targeted manner, help students identify and fill in gaps, and promote the continuous improvement of their abilities and qualities. At the same time, the accumulation of evaluation data also provides a scientific basis for teachers to optimize teaching content and improve teaching methods.

To effectively evaluate the learning effectiveness of non-computer science students in technical courses, this study constructs a three-dimensional evaluation model of "process result development".

(1) Process evaluation (40%) integrates IRS real-time feedback system, non professional version experiment report gauge, and Scrum milestone evaluation [8], focusing on monitoring classroom participation (10%), subject adaptation experiment report (15%), and project phase results (15%). The IRS system quantifies students' ability to transform technical concepts (40%) and the quality of interdisciplinary problem posing (60%) through real-time collection of question/discussion data; Design a non professional scoring scale for the experimental report, highlighting the disciplinary relevance of data interpretation (30%), the rationality of technical tool application (40%), and the professional adaptability of visual presentation (30%).

(2) Consequential evaluation (50%) includes final project (30%), technical defense (15%), and theoretical testing (5%). The final project requires completion of the entire process from business analysis to visualization, with a focus on evaluating the professional applicability of technical methods (50%) and the practical value of analytical conclusions (30%); The defense session adopts a core statement and minute scene Q&A mode, focusing on the accessibility of technical communication (40%). (3) Developmental evaluation (10%) collects multi-source learning behavior data based on the xAPI standard [9], and generates personalized improvement suggestions through analysis.

Based on Gardner's theory of multiple intelligences [10], three types of professional assessment paths are designed: management, which uses financial analysis of listed companies as a carrier to assess Pandas index calculation (40%) and Plotly dashboard decision support (30%); Humanities focus on social public opinion analysis, evaluating the depth of humanistic interpretation in TextBlob sentiment analysis (50%) and the narrative effectiveness of Kepler.gl geographic visualization (30%); Science and engineering majors use environmental monitoring modeling to examine the engineering applicability of the Scikit learn model (45%) and the parameter scientificity of SHAP feature analysis (35%).

4. Construction of Teaching Resources

4.1. Resource Integration Design Based on Existing Platforms

To reduce the cost of local data maintenance and cultivate students' ability to obtain data in real scenarios, this study suggests fully utilizing the infrastructure resources of mature cloud platforms such as Kaggle and Alibaba Cloud. By building a hybrid teaching resource system on a cloud platform, local facility development costs can be saved (compared to self built testing data), and students can directly operate real industry datasets. Build a two-level resource system of "core expansion". The core layer directly calls Kaggle Kernels' interactive case library [11] and Alibaba Cloud DataWorks industry dataset (15 categories including healthcare/education) [12], while the extension layer develops customized micro course videos for professional characteristics, using a three-level content structure of "foundation case extension". All sensitive data is anonymized through Alibaba Cloud's data anonymization service.

4.2. Lightweight Practice Environment Renovation Plan

Based on cloud native service refactoring technology stack: deploying Kaggle interactive environment in programming environment, supporting concurrent access by 200 people; Develop multiple domain specific templates (including data visualization debugging plugins) using the pre built mirror function of Alibaba Cloud DSW; Integrate Kaggle error pattern library and custom rules for error detection.

4.3. Cross Platform Support Tool Development Strategy

Utilize existing platform resources to shorten the construction cycle of the case library and develop three types of bridging tools: cloud platform terminology converter, which maps technical terms to professional terminology; Error solution knowledge base, aggregating Kaggle solutions and Alibaba Cloud help documents for error knowledge base; Learning Kanban integrates multi platform learning data through APIs.

5. Conclusion

This study examines key pedagogical challenges in delivering foundational big data education to non-computer science students , proposing a holistic pedagogical framework to resolve persistent issues such as ambiguous curriculum positioning, insufficient interdisciplinary integration, and gaps in practical competency development. By redefining instructional objectives through the lens of cross-disciplinary literacy, the framework emphasizes the cultivation of data-driven critical thinking tailored to diverse professional domains, while systematically embedding ideological and political education to enhance the curriculum’s ethical and societal relevance.

The instructional methodology employs a blended teaching approach. The curriculum integrates modularized content structured around "foundational concepts–domain applications–ethical considerations," utilizing adaptive case studies ranging from lightweight single-domain exercises to cross-disciplinary collaborative projects. Enterprise partnerships enable the infusion of authentic business scenarios—such as e-commerce user behavior analysis and environmental monitoring—into experiential learning, fostering technical proficiency in data processing pipelines (collection, cleaning, analysis, visualization) alongside teamwork and problem-solving skills.

The three-dimensional "process-outcome-development" evaluation model replaces conventional assessment paradigms by dynamically monitoring learning trajectories. Process-oriented metrics track conceptual comprehension through real-time classroom interactions and milestone-based project evaluations, while outcome assessments prioritize domain-specific technical deliverables, such as financial indicator dashboards for business majors or sentiment analysis reports for humanities students. Development-focused evaluations leverage longitudinal learning analytics to generate personalized feedback, addressing individual competency gaps.

While the framework demonstrates theoretical coherence, two limitations necessitate further exploration. First, dependencies on third-party platforms (e.g., Kaggle, Aliyun) introduce potential risks in data sovereignty and curriculum continuity, requiring hybrid architectures that balance cloud-based resources with localized backup systems. Second, the evolving nature of ethical challenges in data science—particularly privacy preservation and algorithmic accountability—demands ongoing curriculum updates to align with global standards like the GDPR and China’s Data Security Law.

Future enhancements should focus on three strategic directions: 1) Intelligent teaching aids, including AI-driven adaptive learning systems for customized content delivery and automated assessment tools powered by large language models (LLMs), to optimize instructional efficiency; 2) Ecosystem sustainability through graduate employment tracking mechanisms that correlate curriculum outcomes with workplace performance, enabling iterative refinements; and 3) Global-local synergy, blending international best practices (e.g., ACM Data Science Competency Framework) with localized pedagogical traditions to address regional workforce demands.

This reform blueprint transcends disciplinary boundaries, repositioning big data education as a catalyst for digital transformation across academic fields. By interweaving technical mastery, domain contextualization, and ethical awareness, it equips non-computer science disciplines with future-proof data literacy—the capacity not merely to manipulate tools but to critically interrogate and innovate within data-rich environments. The integration of ideological education further ensures that technological empowerment aligns with societal values, cultivating responsible digital citizens poised to navigate the complexities of the algorithm-driven era. As higher education institutions worldwide grapple with digital upskilling imperatives, this framework offers a transferable model for transforming niche technical training into broad-based competence cultivation, ultimately bridging the gap between specialized knowledge silos and the interconnected realities of the digital economy.

References

[1] Mayer-Schönberger, V., & Cukier, Kenneth (2013). Big data:A revolution that will transform how we live, work, and think. Eamon Dolan/Houghton Mifflin Harcourt.

[2] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

[3] Crawley, E. F., Malmqvist, J., Östlund, S., & Brodeur, D. R. (2014). Rethinking engineering education: The CDIO approach (2nd ed.). Springer.

[4] Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy. Pearson.

[5] ACM Data Science Task Force. (2021). ACM data science curriculum guidelines. Association for Computing Machinery.

[6] IBM Research. (2012). Big data analytics: Disruptive technologies for changing the game [Technical white paper]. IBM Corporation.

*[7] Kaggle Team. (2020). *Kaggle datasets documentation: COVID-19 and stock market data* [Technical documentation]. Retrieved from https://www.kaggle.com/docs/datasets*

[8] Schwaber, K., & Sutherland, J. (2020). The scrum guide: The definitive guide to scrum. Scrum.org.

[9] Advanced Distributed Learning Initiative (ADL). (2016). xAPI (experience API) specification. U.S. Department of Defense.

[10] Gardner, H. (2011). Frames of mind: The theory of multiple intelligences (3rd ed.). Basic Books.

[11] Kaggle Team. (2023). Kaggle kernels documentation. Retrieved December 1, 2023, from https://www.kaggle.com/docs/kernels

[12] Alibaba Cloud. (2022). DataWorks technical white paper: Data integration and governance [Technical white paper]. Alibaba Cloud.