

# Constructing and Implementing a Practice-Oriented Teaching System for Data Science and Big Data Technology Based on Competition-Integrated Education

Hexia Cheng<sup>1\*</sup>, Haifeng Wu<sup>1</sup>, Hesheng Cheng<sup>2</sup>

1 School of Artificial Intelligence and Computer Science, Anqing Normal University, Anqing, Anhui 246011, China

2 School of Computer and Artificial Intelligence, Hefei Normal University, Hefei, Anhui 230000, China

---

**Abstract:** In response to the challenges faced by the Data Science and Big Data Technology program in practical teaching—such as insufficient software and hardware resources, delayed updates of instructional software, fragmentation among theoretical courses, and difficulties in quantitatively assessing teaching outcomes—it is imperative to explore a teaching reform pathway that can effectively integrate theoretical instruction with practical application. Taking into account the interdisciplinary nature and strong application orientation of the Data Science and Big Data Technology program, this paper proposes a practice-oriented teaching reform model based on the concept of “competition-driven learning.” By introducing high-level discipline competitions with strong alignment to the major, competition projects are systematically embedded into both theoretical and practical courses, thereby restructuring the organization of course content. Furthermore, competition outcomes and certificates are employed as objective criteria for evaluating the teaching process and learning effectiveness. Practical implementation demonstrates that this teaching model effectively breaks down knowledge barriers between courses, promotes deep integration of theoretical knowledge and engineering practice, and significantly enhances students’ engineering practice capabilities, innovative application skills, and teamwork awareness. Moreover, it provides a replicable and referential implementation framework for reforming talent cultivation models in related disciplines.

**Keywords:** competition-integrated education, academic competitions, practice-oriented teaching system, applied talent development, Data Science and Big Data Technology

---

## 1. Introduction

Data Science and Big Data Technology play a crucial role in both emerging technologies and traditional manufacturing sectors, exerting significant influence on technological innovation and industrial upgrading, among other aspects [1]. Anqing Normal University began enrolling students in the Data Science and Big Data Technology program in 2019. Drawing on nearly five years of teaching practice, the university has carried out a systematic reflection and evaluation of the program’s talent cultivation and teaching processes, promoting effective instructional practices while implementing targeted improvements and refinements to address identified shortcomings. As a well-known provincially administered application-oriented university, Anqing Normal University places considerable emphasis on practical courses in talent cultivation. However, key challenges remain, including how to identify

engineering projects with real application value that are suitable for student practice, how to effectively apply theoretical knowledge to practical scenarios, and how to cultivate students' abilities in project task allocation and collaboration, hands-on practical skills, and problem-solving capabilities in real-world contexts. These issues urgently require systematic consideration and resolution.

With the continuous enhancement of professional competencies, discipline competitions have become a key vehicle for fostering students' innovative capabilities in the field of Data Science [2]. By engaging students in authentic, industry-inspired challenges, competitions shift learning from passive technical execution to active problem framing and resolution. When knowledge gaps arise during competition preparation, students are motivated to independently acquire new skills—thereby fostering self-directed learning and adaptive expertise [3]. Moreover, insights from competition tasks provide instructors with actionable feedback to refine curricula, aligning academic content with industrial demands and closing the gap between education and practice. This dynamic creates a closed-loop “competition-driven learning” ecosystem.

## 2. Characteristics of the DSBDT Discipline

DSBDT is inherently interdisciplinary and application-oriented, exhibiting four key traits:

(1) **Multidisciplinary Integration:** It synthesizes computer science, mathematics, and domain-specific knowledge, using data as input, computational tools for processing, mathematical models for inference, and visualization techniques for communication.

(2) **Engineering-Centric Practice:** Core activities—such as data cleaning, analysis, system development, and decision support—are rooted in real-world projects from industry, competitions, or lab experiments.

(3) **Data as Foundational Resource:** High-quality, large-scale datasets are essential for training AI models (e.g., recommendation systems) and extracting actionable insights; without them, downstream analytics lose validity.

(4) **Rapid Technological Evolution:** The field has progressed from single-machine SQL processing → distributed frameworks (e.g., Hadoop/Spark) → intelligent real-time systems (stream computing + AI integration), emphasizing immediate value extraction from data.

## 3. Identified Gaps in Current DSBDT Education

Despite its applied orientation, DSBDT education faces systemic limitations:

(1) **Inadequate Hardware Resources:** Most university clusters lack sufficient nodes or GPU/TPU capacity to handle TB-scale data or real-time workloads (e.g., graph or stream processing), restricting students to toy datasets.

(2) **Outdated Software Platforms:** Teaching environments often rely on legacy open-source versions (e.g., Hadoop 2.x) without enterprise features like Kerberos authentication or YARN resource isolation. Fewer than 15% support emerging technologies such as vector databases (Milvus) or large model fine-tuning (LoRA) [4].

(3) **Fragmented Knowledge Delivery:** Courses in statistics, programming, and domain applications are taught in isolation, lacking integrated case studies that mirror real-world project complexity.

(4) **Subjective Assessment:** Project-based evaluations rely heavily on final reports, offering little insight into process, effort, or incremental progress—making performance quantification difficult.

Notably, academic competitions directly address these shortcomings by providing industrial-grade computing resources, up-to-date toolchains, cross-disciplinary problem contexts, and standardized evaluation metrics (e.g., leaderboard rankings, official certificates).

## 4. Strategic Selection of Academic Competitions

Drawing from the *2023 National Report on Undergraduate Academic Competitions* (which lists 84 nationally recognized contests), we identify 12 high-impact competitions relevant to DSBDT, including:

- (1) China Collegiate Computing Contest – Big Data Challenge
- (2) National Undergraduate Contest on Big Data & Artificial Intelligence
- (3) “Internet+” Innovation & Entrepreneurship Competition (Big Data Track)

(4) Huawei ICT Competition (Big Data Track)

(5) Blue Bridge Cup (Big Data Category)

(6) National Statistical Modeling Competition

Three exemplars are analyzed in depth:

(1) “Internet+” Innovation & Entrepreneurship Competition (Big Data Track)

Sponsored by Alibaba Cloud and Huawei Cloud, this contest provides TB-scale data processing environments where students apply Hadoop/Spark in real scenarios (e.g., financial risk control, medical analytics). It reinforces coursework in *Big Data Analytics* and *Data Mining* while developing business acumen through cross-institutional teaming and pitch presentations. Top performers gain graduate admission bonuses or internship opportunities at leading tech firms.

(2) Huawei ICT Competition (Big Data Track)

This competition bridges the academia-industry platform gap by offering enterprise-grade MRS 3.0 (Huawei’s big data stack) and access to anonymized production datasets (e.g., banking transactions, IoT time-series logs)—far exceeding typical UCI repository usage. Winners receive Huawei-certified credentials and are added to the “Huawei Talent Pool,” granting visibility to over 6,000 ecosystem partners like iFLYTEK.

(3) China Collegiate Computer Game AI Tournament

This contest uniquely integrates algorithm design, probability theory, and software engineering through game AI development (e.g., Go, Chinese Chess, Poker). Tasks involve Monte Carlo Tree Search (MCTS), heuristic evaluation functions, and real-time decision-making under uncertainty. The 2024 edition introduced LLM-enhanced strategy optimization. Certificates are issued by the Chinese Association for Artificial Intelligence, with top candidates fast-tracked to recruitment pipelines at Alibaba and Huawei.

## **5. Integration of Disciplinary Competitions into Professional Theory and Practice Courses**

High-quality disciplinary competitions can effectively integrate disciplinary knowledge and enhance students' practical abilities <sup>[5]</sup>. In the context of disciplinary competitions, both professional theoretical courses and practical courses should appropriately incorporate cutting-edge applied knowledge. Taking the China University Computer Gaming Competition as an example, the knowledge and skills required for its Gomoku project are distributed across courses like the C programming language and Data Structures, specifically involving areas such as data structures, enumeration, queues, traversal, and probability theory. By decomposing the projects from the game competition into sub-projects and integrating them as experimental components of the Data Structures course, students build foundational skills for the competition. The final competition program synthesizes knowledge acquired throughout the course, thereby strengthening the close relationship between the competition and the curriculum. This approach not only enhances learning efficiency but also improves problem-solving capabilities.

(1) Mapping of Core Competition Knowledge to Course Content

C language provides the underlying support for the engineering implementation of the project; pointers and memory management enable efficient storage of game states, such as the dynamic allocation of a two-dimensional board array. File I/O operations are used for loading opening libraries and recording game logs. Multithreaded programming enables concurrent control for real-time gameplay, achieving millisecond-level response in disciplines such as Texas Hold'em. Data structures serve as the carrier of game logic, fundamental algorithms form the core of decision intelligence, and the Monte Carlo Tree Search (MCTS) algorithm addresses games with uncertain information. The evaluation function is designed by combining heuristic rules (e.g., piece value weights in Junqi) with machine learning models. The optimization strategy employs genetic algorithms to dynamically adjust the parameters of the evaluation function, as demonstrated in practical cases of Chinese Chess.

## (2) Knowledge Integration Through Competitions

Discipline competitions facilitate the integration of scattered knowledge into an organic system. Diagram of Task-Driven Knowledge Integration Across the Entire Process Chain.

Typical Competition Cases Demonstrating Cross-Curricular Collaboration

Case 1: Go AI Development

C Programming: Implementing board state compression and storage using bitwise operations for memory optimization.

Data Structures: Using sparse matrices to store stone placement relationships.

Algorithms: Parallelizing Monte Carlo Tree Search (MCTS) accelerated by integrating the UCB formula.

Case 2: Real-time Decision-Making in Texas Hold'em

Probability Models (Probability Theory Course): Calculating winning odds and balancing risk versus reward.

Stream Computing (Distributed Systems): Processing opponent action streams using Kafka and Flink.

## (3) Pedagogical Value: Compensating for the Deficiencies of Traditional Curricula

Real-world engineering problems inherently require interdisciplinary approaches for solutions. For instance, completing a face recognition project involves multiple stages: firstly, data acquisition, data cleaning, data labeling, and dataset training; secondly, employing artificial intelligence methods for learning to facilitate subsequent software deployment; and finally, deploying the project on appropriate hardware platforms. Such a project integrates knowledge that cannot be achieved within a single discipline.

However, current teaching content often remains divided into basic courses and specialized courses, with limited connection between specialized subjects, leading to the "knowledge silos" problem. A typical example in university teaching is the separation between the C Programming Language course and the Data Structures course. This disconnection prevents students from effectively linking pointer operations with the construction and manipulation of tree structures.

In contrast, competitions compel the development of a complete game system, which necessitates the integration of knowledge across these domains.

Develop computational thinking by practicing four key aspects — decomposition, pattern recognition, abstraction, and algorithm design.

This approach directly combats “knowledge silos.” For instance, traditional separation of C Programming and Data Structures often leaves students unable to link pointer manipulation to tree traversal. Competition mandates force synthesis: building a functional game AI requires seamless integration of memory management, search algorithms, and probabilistic modeling.

## 6. Conclusion

Anqing Normal University adheres to a student-centered, regionally responsive educational mission. By embedding high-caliber academic competitions into the DSBDT curriculum, we have created a dynamic learning environment where students engage with industrial-scale data, modern toolchains, and authentic problem-solving contexts. This competition-integrated model not only strengthens foundational knowledge but also cultivates the practical, innovative, and collaborative competencies demanded by regional industries. Our experience offers a scalable blueprint for applied talent development in data-intensive disciplines worldwide.

## Acknowledgments

Funding: This work was supported by the Provincial Quality Engineering Projects of Anhui Higher Education Institutions (Grant Nos. 2023syjy034, 2023xjzlt042, 2024cywzy026, 2023sdx044).

## References

[1] Du, Y. B., Wu, X. Q., Liu, C. H., et al. (2024). Exploring Elite Talent Cultivation Models for Statistics Majors in the Big Data Era. *Advances in Education*, 14(12), 2367. <https://doi.org/10.12677/ae.2024.14122367>

[2] Gao, Y., Liu, H. F., & Xu, L. L. (2023). Integrating Academic Competitions into Practical Teaching of Data Science and Big Data Technology. *Henan Education (Higher Education)*, (8), 80–82.

[3] Wang, H. Y., Hu, Y. X., & Zhang, W. (2024). Exploration and Practice of Practical Teaching in Big Data Technology under the New Engineering Context. *Education and Teaching Forum*, (18), 93–96.

[4] Wang, X. X., Tan, P., & Gao, S. (2025). Reforming “Algorithm Design and Analysis” Teaching for Data Science Majors under the New Engineering Paradigm. *Advances in Education*, (4), 299–307.

[5] Wu, J., & Deng, H. H. (2024). Competition-Driven Reform in Spark-Based Big Data Technology Teaching. *Frontiers in Educational Research*, (4), 33–36.

### **Author Bio**

Corresponding author: Hexia Cheng (born July 1982), male; Han Chinese; hometown: Anqing, Anhui; educational level: master's degree; title: associate professor; research direction: computer applications; E-mail: chenghexia@aqnu.edu.cn

Haifeng Wu (born August 1982), male; Han Chinese; hometown: Anqing, Anhui; educational level: master's degree; title: professor; research direction: computer applications.

Hesheng Cheng (born January 1985), male; Han Chinese; hometown: Anqing, Anhui; educational level: master's degree; title: lecturer; research direction: computer applications.