Jinan University
Undergraduate Syllabus for Educational Data Analysis and Mining (Course Title)

Course code 60080080

Course title Educational Data Analysis and Mining

Course type Compulsory [ ] Specialized Optional [√] Common Optional [ ]

Prerequisites NA

Credits 3

Total course hours 54

Majors applicable to

Students classification Mainland [√] Non-mainland [√]

Department/School offering the course College of Information Science and Technology

I. Teaching Objectives and Requirements

This course aims to prepare students with fundamental knowledge and skills necessary to perform educational data analysis and mining, to interpret and evaluate methodology and outcomes, and to apply appropriate analytical strategies to address questions that arise in educational research and practice.

II. Key Points and Main Areas of Difficulty of the Course

This course is designed to provide an overview of the fundamental topics and issues in educational data (e.g., test data) analysis and mining. The goal of the course is to offer basic knowledge and techniques required to analyze educational data from a measurement and technical perspective. The topics include (a) principles of educational measurement, such as scaling, reliability and validity, item analysis, classical test theory (CTT), item response theory (IRT), and principles of test construction, and (b) principles of computer science and deep learning, such as machine learning algorithms, and their applications in educational data contexts.

III. Lab or Practical Work Required to Support the Course

1) Data Preprocessing Workshops: Students will learn to clean, preprocess, and transform educational data sets using tools like Python and R. These sessions will focus on techniques for dealing with missing data, noise, or feature selection, which are common issues in educational datasets.
2) Classification and Clustering Labs: Using popular data mining software and libraries (e.g., scikit-learn, Weka), students will apply classification and clustering algorithms to educational data to identify patterns and predict educational outcomes. These labs will help students understand the application of these methods in segmenting student populations and predicting their performance.
3) Association Rule Mining Sessions: Practical exercises on association rule mining will be conducted, where students will use algorithms to discover interesting relationships between variables in large databases of educational content and student interactions.

IV. Textbooks and References

Textbook:

1. Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory*. Waveland Press.
2. Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.

Reference Materials:

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and*

psychological testing. American Educational Research Association.
2. Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Holt, Rinehart and Winston.
3. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
4. Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques*. Morgan kaufmann.

V. Grading System and Evaluation Methods

Participation and Homework 30% + Mid-term exam 20% + Final project 50% = 100%

1. Participation and Homework

Various in-class learning activities and homework assignment will be given, aiming to facilitate and enhance your learning by reflecting on the progress of your learning and reviewing how effectively you have learned to identify strengths, gaps and misconceptions in your knowledge and understanding. The in-class participation and homework assignment collectively worth 30% of the final grade. 20% for homework assignments and 10% for seminar reflections. Note that you need to attend at least 3 seminars hold by Guangdong Institute of Smart Education, online or offline (at the main campus), and write seminar reflection for each attended seminar (each about 200-400 words).

2. Mid-term exam

A mid-term exam comprising multiple-choice and constructed-response questions will be given at Week 9, on the focus of students' understanding of the logic, concepts and applications of the concepts and techniques instead of memorization. The mid-term exam is worth 20% of the final grade.

Since we do not have time for a mid-term exam, I decide to change this part to a lecture practice. Students will be divided into six groups (4 or 5 students per group) and choose a topic. This is still worth 20% of the final grade.

Topics:

Week 2	Classical test theory
Week 3	Reliability
Week 4	Validity
Week 5	Test construction
Week 6	Item analysis
Week 7	Item response theory

For the presenting group:

1. Read and study the materials that I provided to prepare for the lecture.

2. Deliver a high-quality lecture.
3. Answer and discuss students' questions (and my questions).
4. Optional: You can also prepare some questions for students to do group discussion in the class.

For students:

1. You are recommended to read the materials prior to the class.
2. Prepare questions to the presenting group. Discuss and share your thoughts. Each of the five groups should raise at least one question.
3. Write down what you learned from the class and evaluate how well the presenting group has done. All students in the class need to do this, including the presenting students.

3. Final Project

A final project is designed for students to apply the knowledge and techniques learned in this course (LLMs are recommended) to analyze empirical data and address specific educational data questions.

The final project is conducted in a 3-person team. Each team is required to (1) submit a 5-min pre-recorded video about the final project; (2) give an oral presentation in the last class (Week 16; 10 minutes) and (3) write a final project report. The final project is worth 50% of the final grade.

VI. Detailed Points of Teaching Contents

54 teaching hours include 48 teaching hours for lectures and 6 teaching hours for seminars.

Students are required to attend at least 3 seminars (2 teaching hours for each) about educational data analysis and mining.

Week 1 (9/3)	Introduction to educational measurement and basic statistics
Week 2 (9/10)	Classical test theory
Week 3 (9/17)	Reliability
Week 4 (9/24)	Validity
Week 5 (10/1)	Test construction
Week 6 (10/8)	Item analysis
Week 7 (10/15)	Item response theory
Week 8 (10/22)	Educational testing and measurement (optional; self-study recommended)
Week 9 (10/29)	Mid-term test
Week 10 (11/4)	Introduction to educational data mining
Week 11 (11/11)	Classification techniques
Week 12 (11/18)	Clustering techniques
Week 13 (11/25)	Deep learning and neural networks
Week 14 (12/2)	Multimodal machine learning in education
Week 15 (12/9)	Large language models in education
Week 16 (12/16)	Student presentation & Course summary

Author's signature_____ Examiner's signature_____

Date _____ Date_____

(dd/mm/yy) (dd/mm/yy)

Instructions

I. This form should be completed after discussion between the Teaching and Research Section staff of the department concerned, and then confirmed and sealed by the corresponding college. Regarding any specialized Compulsory Courses and specialized optional courses, which are not attached to any specific Teaching and Research Section, or are lectured by one teacher alone, discussion should be organized by the department, and confirmation signed and sealed by the college. In the case of Common Optional Courses, syllabuses should first be drawn up by the lecturer in charge, then demonstrated and confirmed by specialists selected by the Academic Affairs Department (AAD) of the university, and finally approved and signed by the Director in charge.

II. The Contents of the Syllabus should be consistent with Undergraduate Program of Year 2006 Students.

III. "Course Type" refers to Compulsory Specialized Optional or Common Optional course.

IV. "Students Classification" refers to Mainland Student or Non-Mainland Student.

V. "Majors Applicable to" refers to all the major titles that could be applied to a certain syllabus. "Public Compulsory Course" should be marked as "All Majors of the University".

VI. "Course offered by" refers to college, Department and Teaching and Research Section, or simply College and Department if the course has no Teaching and Research Section.

VII. "Total Course Hours" are counted as 18 hours per I credit for one course.

VIII. "Practice-oriented Teaching Contents and Requirements". Name of practicing/internship to be noted. For detailed information, see "Laboratory Syllabus" or "Internship Syllabus".

IX. "Textbooks and References" refers to the main textbooks used. Instructors should select textbooks of high quality, or published within 3 years of the current academic year; or those recommended by specified research fields. If the instructors choose self-edited textbooks/ references, they must document the titles of the textbooks or references and state the reasons for their choice.

X. "Grading System and Evaluation Methods". Grading System refers to the percentage allocation of final scores, with assignments, quizzes and attendance mark etc not exceeding 40% of the total score; Evaluation Methods refers to the form of the examination, e.g. "open-book", "closed-book" or "thesis writing" and so on.

XI. "Main Points of Teaching Contents". Instructors should fill in the objectives and requirements, key points, class hour allocation of each chapter, as well as the main contents of each chapter.