Capital University of Economics and Business
Overseas Chinese College
Course Syllabus

Year and Semester	2021 Fall(senior)
Course Name	Big Data
Course Code	MIS403
Course Type	☐ General Education (Required) ☐ General Education (Elective) ☐ Professional Course (Required) ☑ Professional Course (Elective) ☐ Basic Disciplinary Course
Course Credits	3
Course Hours	48
Prerequisites	Statistics, Database, SQL
Instructor	Leilei Zhu (Emma Zhu)
Contact Information	Office: C217 Tele: 15801473268 Email: zhuleilei@cueb.edu.cn
Office Hour	M: 15:30-17:30, W: 8:00-9:00, 11:00-12:00, TH: 8:00-9:00, 11:00-12:00
Learning Centre	M: 18:00-20:00, T: 10:00-12:00
Grade/Section	2018IT
Course Time/Place	T 13:30-16:30 /B208

Textbook

1. Viktor Mayer-Schönberger. *Big Data:A Revolution That Will Transform How We Live, Work, and Think, First Edition*. Zhejiang Renmin Press, ISBN: 978- 7- 213-05254-5
2. Robert l. Kabacoff, *R In Action Data analysis and graphics with R, Second Edition.* Manning Publications, ISBN: 978-1-617-29138-8

Reference Book

1. Viktor Mayer-Schönberger. *Delete: The Virtue of Forgetting in the Digital Age, First Edition*. Zhejiang Renmin Press, ISBN: 978- 7- 213-05251-4.

Course Description

Big data is an introductory course for students majored in information system management. The main content is basic introduction to big data and algorithms, topics covered will include concept and features of big data, Hadoop system, regression, clustering and classification algorithms as well as cases of applications. Students will not only develop skills of data analysis and ability of data-driven decision making, but also scientific thinking which are all indispensable for future study and professions.

Student Learning Objectives

After completing this course, students will be able to:

Knowledge:

- describe concept and features of big data;
- describe correlation, datafication, risks of big data time;
- identify functions of blocks of Hadoop ecosystem;
- describe regression, clustering and classification algorithms;

Capability:

- analyze data using proper algorithms;
- develop ability of data-driven decision making by big data methods;
- construct the scientific thinking and mindset, include systematic thinking, logic thinking, critical thinking and strategic thinking;

Value:

- develop the quality and morals of being objective, integrity and dedication;
- criticize the world with data-driven philosophical view.

Website Source

https://www.webopedia.com/TERM/B/big_data.html

https://www.r-project.org/

https://www.rstudio.com/

Teaching Methods

This course consists of lectures, discussions and student presentations. Students will be divided into small groups with a group leader helping others in the group. Students must be prepared to finish some small questions and small quizzes during the class.

Grade Criterion

Component	Weight	Description
Final Exam	20%	A cumulative final examination will be given based on all of the contents of the class. The exam paper may be composed of multiple-choice questions, short answer questions, essay questions, problems, and preparation of financial statements. Students should rely primarily on homework assignments to give them a sense of what they may see for material on exams.
Mid-Term Test	20%	A cumulative midterm test will be given based on all of the contents that have been taught in class. The test paper may be mainly composed of multiple-choice questions and it should be completed within 15 minutes in class.
Homework	15%	Most of the assigned homework is taken from the Exercises in the textbook. Assignments will be collected at the clearly stated date. Late assignments will not be accepted. The graded assignments will be kept

		by the tutor for reference and won't be returned to students.
Quizzes	15%	There will be at least 2 quizzes during the semester. Quizzes may or may not be announced in advance. It may also be used as a way to check the attendance. Quizzes will test your knowledge of both concepts and the application of those concepts.
Presentation	10%	The students will be divided into several groups to prepare a presentation. Each student is required to be involved in the presentation. The topics can be selected from the textbook or lectures. Each group need to finish a PPT related to the topic which is given and hand in the related resources to the teacher before the presentation.
Participation	10%	Individuals will be asked to participate individually in a question and answer at least 5 times during the semester. The performances should be counted in their participation.
Attendance	10%	Refer to attendance policy listed below
Total	100%	

Detailed Grade Computation

	Before Midterm	After Midterm
Attendance	5%	5%
Participation	5%	5%
Homework	5%	10%
Quizzes	5%	10%
Presentation		10%
Midterm test	20%	
Final exam		20%
Total	40%	60%

Grading Policy

A+ 97-100	A 93-96	A- 90–92	B+ 87-89	B 83-86	B- 80–82
C+ 75-79	C 70-74	C- 67–69	D+ 63–66	D 62-60	F 0- 59

Exam Schedule

Midterm Test: 8th week

Final Exam:

Assessment of Student Performance

☞ ***Self-Study and Reading ability Practice***

Instructor will give out the chapters or the reference books to read and use class hours to have discussion; students should be able to show a proactive attitude and ability for self-study and reading. Knowledge and oral English will be elements of homework or presentation score.

☞ ***Homework***

Students should finish their homework by themselves. Copying from others will be treated as cheating and the homework scores will be lowered. Students should hand in all assignments on time. Late

assignments will be accepted at the discretion of the instructor (i.e., when the student was ill or had an excused absence). Late assignments without reasonable proof will be reduced in score by 50%.

☞ ***Attendance***

Because the course covers a great deal of material, attending every class session is very important for performing well.

- Being late for 15 minutes or more is considered an absence.
- Five hours or above of unexcused absences will result in the lower level of the final grade by one grade band (e.g. from C – to D +). Any excused absence must be discussed directly with the teacher.
- Absence which is more than 1/3 of the total teaching hours will cause an F (a failing grade) directly.

but students are welcome to continue attending classes.

- An incomplete grade (I) will be considered in case of medical or family emergencies.

☞ ***Participation***

- Students should participate in classes actively. Half of participation grade is determined by their presentation in class. They are encouraged to ask questions relevant to the subject and express their own opinions. Every student should respect the ideas, opinions, and questions of their classmates.
- Students should also use office hours to ask questions or talk with the instructor for good communication and effective learning.
- Frequent visiting the instructor and chatting in English during office hours is highly recommended.
- Any misbehavior and non-class related activities in class will result in the lower level of the participation grade, including ringing cell phones.
- All above behaviors will be solely evaluated by the instructor for scoring.

☞ ***Textbook***

Students must bring the textbook to class.

Topical Course Outline

Week	Date	Content	Course Education
1		**Syllabus & Orientation** **Chapter 1 Brief Introduction of Big Data** Section 1.1: Basic concepts Section 1.2: Official definition of big data Section 1.3: Features of big data	
2		**Chapter 1 Brief Introduction of Big Data** Section 1.4: Challenges and opportunities Section 1.5: Applications of big data analysis	Critical thinking
3		**Chapter 2 Brief Introduction of Hadoop** Section 2.1: What is Hadoop? Section 2.2: Why Hadoop?	Systematic thinking

		Section 2.3: The Hadoop ecosystem	
4		**Chapter 2 Brief Introduction of Hadoop** Section 2.4: Introduction of HDFS Section 2.5: Features of HDFS Section 2.6: How HDFS works Section 2.7: Introduction of MapReduce	
5		**Chapter 2 Brief Introduction of Hadoop** Section 2.8: What is Hive? Section 2.9: Applications of Hive Summary	
6		**Chapter 3 Correlation (Self-study and presentations)** Section 3.1: Predictions and Predilections Section 3.2: Illusions and Illuminations Section 3.3: Man and Manhole Section 3.4: The End of Theory?	logic thinking and critical thinking
7		**Chapter 4 Datafication (Self-study and presentations)** Section 4.1: Qualifying the World Section 4.2: When Words, locations and interactions become data Section 4.3: The Datafication of Everything Section 4.4: More examples and cases	
8		Midterm test	
9		**Chapter 5 Introduction of R** Section 5.1: Introduction of R and rstudio Section 5.2: R objects and functions Section 5.3: Other information of R	Data analysis skills
10		**Chapter 6 Data Visualization** Section 6.1: Introduction of ggplot2 Section 6.2: qplot Section 6.3: ggplot	Data visualization skills
11		Practice of programming	
12		**Chapter 7 Regression** Section 7.1: Regression with **lm** Section 7.2: Several regression models in R Section 7.3: Regression diagnostics	logic thinking, critical thinking and data analysis skills
13		**Chapter 8 Clustering** Section 8.1: Common steps in cluster analysis Section 8.2: Hierarchical cluster analysis Section 8.3: Partioning cluster analysis	logic thinking, critical thinking and data analysis skills
14		**Chapter 9 Classification** Section 9.1: Classifying with logistic regression Section 9.2: Classifying with decision trees	logic thinking, critical thinking and data

		Section 9.3: Classifying with random forest Section 9.4: Classifying with vector machine	analysis skills
15		Practice of programming	
16		Presentations	
17		Review	
18&19		Final examination	

Note: *Some chapters or sections may leave for self-study, this is the students' duty to learn and understand, they may also be included in the quizzes or exams Final exam is in term of presentations.*
A review in Chinese may be held during L.C. and O.H. in the semester.

Teacher's Office Hour

- The instructor's office hour is shown in the front of the office door.
- Students are suggested to use the instructor's office hour and learning center to ask questions or talk with the instructor once at least per week for good communication and effective learning, which is recorded in the students' participation.
- The time can be scheduled by instructors or students, or both.

Cheating and Plagiarism

Cheating is not tolerated. Any student caught cheating on a quiz; test or exam will be given a mark of zero (0) for the particular work. At the beginning of the semester the definition of plagiarism will be carefully explained, when any thoughts or writings of another person are used, they must be clearly identified (usually one uses quotation marks) and the source notes. **If any student is caught cheating on any homework assignment, the highest score the student can earn in that course is a "C".**

Important Dates

Sep. 5	Registration (Sophomores, Juniors and Seniors)
Sep. 6	Classes Begin (Sophomores, Juniors and Seniors)
Sep. 10	Last Day to Drop or Add a Course (Sophomores, Juniors and Seniors)
Sep. 18	Registration (Freshmen)
Sep. 20-24	Entrance Education (Freshmen)
Sep. 21	Mid-Autumn Festival
Sep. 27	Classes Begin (Freshmen)
Oct. 1	National Day
Nov. 1-5	Midterm Test
Jan. 1, 2022	New Year's Day
Jan. 1-4	Revision (Sophomores, Juniors and Seniors)
Jan. 5-14	Final Exam Period (Sophomores, Juniors and Seniors)
Jan. 10-14	Final Exam Period (Freshmen)
Jan. 17	Winter Vacation Begins

Note: *This syllabus is tentative and may be changed or modified throughout the semester. All students will be notified and a new syllabus will be given.*

Instructor: **Emma Zhu** **Department Head:** **Jingning Li**