

# Capital University of Economics and Business

## Overseas Chinese College

### Course Syllabus

|                                   |                                                                                                                                                                                                                                                                                                 |
|-----------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><u>Year and Semester</u></b>   | 2019 Spring (February 25, 2019 - June 21, 2019)                                                                                                                                                                                                                                                 |
| <b><u>Course Name</u></b>         | Introduction to Data Science                                                                                                                                                                                                                                                                    |
| <b><u>Course Code</u></b>         | MIS453                                                                                                                                                                                                                                                                                          |
| <b><u>Course Type</u></b>         | <input type="checkbox"/> General Education (Required) <input type="checkbox"/> General Education (Elective)<br><input type="checkbox"/> Professional Course (Required) <input checked="" type="checkbox"/> Professional Course (Elective)<br><input type="checkbox"/> Basic Disciplinary Course |
| <b><u>Course Credits</u></b>      | 3                                                                                                                                                                                                                                                                                               |
| <b><u>Course Hours</u></b>        | 48                                                                                                                                                                                                                                                                                              |
| <b><u>Prerequisites</u></b>       | Calculus, Probability and Statistics                                                                                                                                                                                                                                                            |
| <b><u>Instructor</u></b>          | Leilei Zhu (Emma Zhu)                                                                                                                                                                                                                                                                           |
| <b><u>Contact Information</u></b> | Office: C217<br>Tele: None<br>Email: zhuleilei@cueb.edu.cn                                                                                                                                                                                                                                      |
| <b><u>Office Hour</u></b>         | TBA                                                                                                                                                                                                                                                                                             |
| <b><u>Learning Centre</u></b>     | TBA                                                                                                                                                                                                                                                                                             |
| <b><u>Grade/Section</u></b>       | 2016BA/Y01                                                                                                                                                                                                                                                                                      |
| <b><u>Course Time/Place</u></b>   | T: 15:40—17:30/F:11:10-12:00 / A204;                                                                                                                                                                                                                                                            |

#### **Textbook**

Cathy O'Neil, Rachel Schutt. *Doing Data Science, Straight Talk from the Frontline*. O'Reilly Media, Inc., ISBN: 978- 1- 449-35865-5

#### **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**

This course will focus on the various aspects of data science. It contains: a general introduction of the concept of data science, some statistical concepts and models which include statistical inference, exploratory data analysis and the data science process, what is algorithms and three basic algorithms, the Bayes Laws and applications of Bayes Laws in data science, Logistics Regression and applications of this regression model in classification, an introduction of Time Stamps data and financial modeling, the concept of data visualization and how data visualization is used to detect fraud, social networks analysis and the concept of data journalism; each topic will be followed by a case of real data analysis.

### **Student Learning Objectives**

After completing this course, students will be able to:

- ♦ know the basic concepts and terminology in data science
- ♦ Gain a full skill of doing exploratory data analysis
- ♦ be aware the basic data science process
- ♦ do complex data analysis using proper statistical models
- ♦ develop a data based mind

### **Website Source**

<http://hadoop.apache.org/>

### **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 |
| <b>Total</b> | 100% |                                         |

### **Detailed Grade Computation**

|               | <b>Before Midterm</b> | <b>After Midterm</b> |
|---------------|-----------------------|----------------------|
| Attendance    | 5%                    | 5%                   |
| Participation | 5%                    | 5%                   |
| Homework      | 5%                    | 10%                  |
| Quizzes       | 5%                    | 10%                  |
| Presentation  |                       | 10%                  |
| Midterm test  | 20%                   |                      |
| Final exam    |                       | 20%                  |
| <b>Total</b>  | <b>40%</b>            | <b>60%</b>           |

### **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: the 8<sup>th</sup> week

Final Exam: June 17-June 21, 2019

### **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**

| Date      | Week Index | Topics                                                                                                                                                                                                                         |
|-----------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0225-0301 | Week 1     | <b>Syllabus &amp; Orientation</b><br><b>Chapter 1 Introduction: What Is Data Science</b><br>What is data science?<br>What are data science jobs?<br>A data science Profile?                                                    |
| 0304-0308 | Week 2     | <b>Chapter 2 Statistical Inference, Exploratory Data Analysis, and Data Science Process</b><br>Basic concepts of probability and statistics<br>Exploratory data analysis<br>The data science process<br>Case study: RealDirect |
| 0311-0315 | Week 3     | <b>Chapter 3 Algorithms</b><br>The concept of algorithms<br>Machine learning algorithms<br>Three basic algorithms: Linear regression                                                                                           |
| 0318—0322 | Week 4     | <b>Chapter 3 Algorithms</b><br>Three basic algorithms: K-nearest Neighbors(K-NN)<br>Three basic algorithms: K-means<br>Thought experiment: automated statistician                                                              |
| 0325-0329 | Week 5     | <b>Chapter 4 Spam Filters, Naïve Bayes and Wrangling</b><br>Introduction about Bayes Laws<br>Case study: spam filter for individual words<br>Spam filter that combines words: Naïve Bayes                                      |

|            |         |                                                                                                                                                                            |
|------------|---------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0401-0405  | Week 6  | <b>Chapter 4 Spam Filters, Naïve Bayes and Wrangling</b><br>Sample R code for dealing with NYT API<br><b>Chapter 5 Logistic Regression</b><br>What is logistic regression? |
| 0408-04012 | Week 7  | <b>Chapter 5 Logistic Regression</b><br>M6D Logistic Regression Case Study<br>Click models and the underlying math<br>Sample R code                                        |
| 0415-0419  | Week 8  | <b>Mid-Term Test</b><br>Case study presentations about data analysis (Data will be collected by each students, and the data will be about themselves)                      |
| 0422-0426  | Week 9  | <b>Chapter 6 Time Stamps and Financial Modeling</b><br>What is Timestamps data?<br>Exploratory data analysis of timestamps data                                            |
| 0429-0503  | Week 10 | <b>Chapter 6 Time Stamps and Financial Modeling</b><br>Introduction about financial modeling<br>The process of financial modeling<br>Evaluation of model                   |
| 0506-0510  | Week 11 | <b>Chapter 9 Data Visualization and Fraud Detection</b><br>Introduction about data visualization<br>Several cases about data visualization                                 |
| 0513-0517  | Week 12 | <b>Chapter 9 Data Visualization and Fraud Detection</b><br>Introduction about data visualization<br>Data visualization at squares                                          |
| 0520-0524  | Week 13 | <b>Chapter 10 Social Networks and Data Journalism</b><br>Social network analysis at Morning Analytics<br>Case-attribute data versus social network                         |
| 0527-0531  | Week 14 | <b>Chapter 10 Social Networks and Data Journalism</b><br>Social network analysis<br>Terminology form social networks                                                       |
| 0603-0607  | Week 15 | <b>Chapter 10 Social Networks and Data Journalism</b><br>Morningside analytics<br>More examples                                                                            |
| 0610-0614  | Week 16 | Quiz and Presentation                                                                                                                                                      |
| 0617-0621  | Week 17 | <b>Final Exam</b>                                                                                                                                                          |

**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**

Spring Semester, 2019

February 25, 2019— July 14, 2019

|                  |                                                |
|------------------|------------------------------------------------|
| Feb 24           | Registration                                   |
| Feb 25           | Classes Begin                                  |
| Apr. 5           | Qingming Festival (tentative)                  |
| Apr. 19          | Spring Sports (tentative)                      |
| May 1            | Labor Day Holiday (tentative)                  |
| June 7           | Duanwu Festival (tentative)                    |
| June 17-21       | Sophomore and Junior students' Final Exam      |
| June 24- July 14 | Sophomore and Junior students' Social Practice |
| June 29-July 7   | Revision and Final Exam Period                 |
| July 8-July 12   | Freshmen's Final Exam                          |
| July 15          | Summer 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