



Haas School of Business
University of California, Berkeley

Syllabus for MBA/EWMBA 247-11
Descriptive & Predictive Data Mining

Fall 2021

Course Overview

The importance of data cannot be overemphasized. With the recent explosion in large-scale, high-resolution data, managers are increasingly required to make data-driven decisions, rather than rely on experience and intuition alone. A key step in data-driven decision making is data mining, which enables businesses to extract useful information and valuable insights from otherwise massive and unorganized data. This hands-on, introductory course teaches you the fundamental principles and practical skills you need to develop data mining models that describe and predict the effects of business decisions in a variety of real-world business applications, including (but not limited to) marketing, finance, and operations. Descriptive models provide insights into "What did happen?" and predictive models, "What could happen?". Topics to be introduced include:

- Exploratory data analysis: frequency distributions, outliers, cross-tabulation, data visualization
- descriptive data mining: k-means clustering, association rules
- predictive data mining: binary classification, confusion matrices, ROC charts, data partitioning for evaluation, logistic regression, k-nearest-neighbors, decision trees

Class Meetings

Day	Dates	Time	Room
Sunday	November 7 & November 21, 2021	9:00am-5:00pm PT	N300 Chou Hall

Instructor Luyi Yang
luyiyang@haas.berkeley.edu
(see bCourses for office hours)

Assistant Daniel Furman
daniel_furman@berkeley.edu
(see bCourses for office hours)

Prerequisite 200S Data and Decisions

Textbook Selected chapters of [*Business Analytics*](#) (4th edition) by Camm et al.
Publisher: Cengage Learning

Software

This course will use R/Rattle and Microsoft Excel (Excel 2016 or Excel 2019 is strongly recommended).

R is a free software environment for statistical computing. [Rattle](#) is a graphical user interface (GUI) for data mining using R, also free and open-source. Office365, which includes Excel, is provided free of charge to all registered Berkeley students. See the campus [software catalog](#). More software installation instructions will be provided on bCourses.

Course Schedule

Session #	Topic	Reading
1 (Sunday, Nov 7)	Exploratory Data Analysis Frequency Distributions, Histograms Summary Statistics, IQR Outliers, Empirical Rule, Boxplot Cross Tabulation, PivotTable, PivotChart	Camm Ch 1, 2, 3
	Descriptive Data Mining k-Means Clustering Association Rules	Camm Ch 5.1 Camm Ch 5.2
2 (Sunday, Nov 21)	Predictive Data Mining Binary Classification Data Partitioning, Overfitting Confusion Matrix Logistic Regression k-Nearest-Neighbors Classification Trees	Camm Ch 9.1, 2 Camm Ch 9.3 Camm Ch 9.4 Camm Ch 9.5

Performance Evaluation

Final course grade is determined by students' total course score relative to their peers, in accordance with the official Haas grading policy for elective courses. The total course score is calculated as follows:

Item	Due Date	Percentage
Class Participation	Sunday, Nov 7 & Sunday, Nov 21	10%
Group Assignment #1	Thursday, Nov 18	20%
Individual Assignment #1	Thursday, Nov 18	5%
Group Assignment #2	Thursday, Dec 2	20%
Individual Assignment #2	Thursday, Dec 2	5%
Take-Home Final Exam	Sunday, Dec 5	40%

Class Participation: The class participation grade will depend on the student's contribution to group learning, which can take the form of good questions, answers, observations, and/or shared experiences. It is not how many times you raise your hand, but how substantive your contributions are. Class attendance is required and do expect "cold calls."

Group Assignments: In group assignments, you will analyze (large-scale) data sets and solve data-driven problems using software tools. You must work in groups of at most 4 people. Conscientious contributions to the group assignments are strictly required. In order to enforce this policy, I will solicit feedback from each of you about the contributions of your team members to group work. Failure to contribute adequately to the group will result in a grade penalty in this course. In addition, a pattern of such behavior will result in an academic performance review.

Individual Assignments: Individual assignments mostly consist of concept-check problems that reinforce your conceptual understanding of the data mining methods taught in class.

Take-Home Final Exam: You will take an online exam to demonstrate your knowledge of data mining. [Honorlock](#) may be used to remotely proctor the exam.

Accommodation for Disabled Students Program (DSP) Students

Students with disabilities seeking academic accommodations in this course need to promptly notify the instructor by providing a Letter of Accommodation from the Disabled Students Program.

Grade Dispute Policy

Any request for regrading must be made in writing to the GSI within 1 week of the return of the assignment/exam. Otherwise, the assigned points are final.

Academic Integrity

UC Berkeley Honor Code: As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others.

The Haas School of Business has a zero-tolerance policy for academic dishonesty. Please review the EW MBA Program's academic integrity [guidelines](#).

Instructor Bio

Luyi YANG is an assistant professor in the Operations and Information Technology Management Group at the University of California, Berkeley's Haas School of Business. His research interests include service operations, business model innovation, digital marketplaces, sustainable operations, and operations-marketing interface. He has taught courses in business analytics, data mining, and operations management. Prior to joining Berkeley Haas, he was an assistant professor of operations management and business analytics at Johns Hopkins University's Carey Business School. He received his PhD and MBA from the University of Chicago, Booth School of Business, and his BS in Industrial Engineering and BA in English, both from Tsinghua University.