

**ISE 535 Data Mining**

Fall 2021

Location: GFS 101**Instructor:** Bruce Wilcox**Office:** GER 216**Office Hours:** TBA**Contact Info:** brucewil@usc.edu**Teaching Assistant:** TBA**Office:** N/A**Office Hours:** TBA**Contact Info:** TBA**Catalog Description**

A course in data preprocessing, data cleaning, data summarization, data visualization, and predictive modeling for classification and regression. Modeling dependencies using Association rules.

Course Description

Data mining is the discipline of extracting useful insights from large quantities of data.

This course is about data analytics tools, methods, and applications to accomplish that purpose. It focusses on data preprocessing, data wrangling, and data cleaning of dataframes. It introduces data visualization techniques to find useful information from preprocessed data. It also shows how to perform the four prominent data mining techniques of classification, clustering, associate rule analysis, and anomaly detection.

The course shows how to apply the afore mentioned methods by means of case studies for model construction and evaluation.

Learning Objectives and Outcomes

- Develop an advanced level of proficiency with the preprocessing, visualization, and statistical analysis of data as well as four of the primary data mining algorithmic techniques: classification, clustering, association analysis, and anomaly detection.
- Develop skills in using the R programming environment and some of its packages that are popularly used in industry by data scientists.
- Review and re-enforce basic statistical concepts that are important in the field of data science.
- Explore the data mining disciplines using business-oriented case studies with “mini-projects”
- *Note: this class deals with data mining techniques associated with traditional, structured (“rectangular”) data. It does not address free text mining or image/video mining as those are specialties with their own classes.*

Prerequisite(s): An undergraduate course on statistics.

Recommended Preparation: ISE 225 (Engineering Statistics I) or equivalent, working knowledge of a programming language.

Course Notes

The course material is available on Blackboard.

Technological Proficiency and Hardware/Software Required

Required software: A programming language is required.

Textbooks

The theoretical material in the course is drawn from the following texts:

- Scmueli, et. al., *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R* (DMBA)
- James, et. al., *An Introduction to Statistical Learning with Applications in R*, Springer, 2017 (ISLR)
- Bruce, et. al., *Practical Statistics for Data Scientists*, O’Reilly, 2020 (PSDS)

In addition, the following texts will be used as references for R programming:

- Wickham, *R for Data Science*, O’Reilly, 2017 (RDS)
- Xie, et. al., *R Markdown: The Definitive Guide*, CRC Press, 2018 (RMD)

Description and Assessment of Assignments

- **Eight homework assignments (one per module)** - 50% of final grade
- **Mid-term exam** – 10% of final grade (covering Modules 1 – 4)
- **Final course project** - 20% of final grade
- **Final exam** - 20% of final grade

Grading Scale (Course final grades will be determined using the following scale)

A	95-100	B-	80-82	D+	67-69
A-	90-94	C+	77-79	D	63-66
B+	87-89	C	73-76	D-	60-62
B	83-86	C-	70-72	F	59 and below

Class Participation: Up to 2 additional points on the overall course grade may be awarded for students based on active class participation.

Assignment Submission Policy

Assignments will all be prepared and submitted using R Markdown unless otherwise directed. They should be submitted via backboard by the due date. Email submissions and late submissions are not allowed.

Timeline and Rules for submission

Assignments are to be returned the week after submission. Solutions will be released soon after the homework submission date.

Course Schedule: A Weekly Breakdown

Week	Date	Topics/Daily Activities	Assignments	References
1	8/27	Module 1: Introduction Introduction to R, RStudio, and R Markdown.	Module 1 HW Assigned	Class notes RMD DMBA Ch. 1&2
2	9/3	Module 2: Data Preparation Data integration, cleaning, reduction, enhancement <i>Tools: Tidy Data, Tidyverse, DPLYR, Tibble</i>	Module 1 HW Due Module 2 HW Assigned	RDS, Section 3 ("Wrangle")
3	9/10	Module 3: Exploratory Data Analysis (EDA) Univariate/bivariate analysis, data quality assessment <i>Tools: ggplot</i>	Module 2 HW Due Module 3 HW Assigned	PSDS, Ch 1 RDS, Section 2 ("Explore") DMBA, Ch 3
4	9/17	Module 4: Statistical Data Analysis Data and sampling distributions, statistical experiments, significance testing	Module 3 HW Due	ISLR, Ch 2 PSDS, Ch 2 & 3
5	9/24		Module 4 HW Assigned	
6	10/1	Mid-Term Exam (90 minutes) Module 5: Classification Logistic regression, linear discriminant analysis, and tree-based methods	Module 4 HW Due Final Project Assigned	ISLR, Ch 4 & 8 PSDS, Ch 5 & 6 DMBA, Ch 7-10
7	10/8		Module 5 HW Assigned	
8	10/22	Module 6: Unsupervised Learning (Clustering) Principal components analysis, clustering (K-means, hierarchical, and model-based)	Module 5 HW Due	ISLR, Ch 10, PSDS, Ch 7 DMBA, Ch 15
9	10/29		Module 6 HW Assigned	
10	11/5	Project Plan Discussion and Case Studies	Project plan due Module 6 HW Due	
11	11/12	Module 7: Association Rule Analysis	Module 7 HW Assigned	DMBA, Ch 14
12	11/19	Module 8: Anomaly Detection	Module 7 HW Due	TBA
13	12/3	Final Exam Review	Final project Due	
Final	12/13	Final Exam (2:00PM – 4:00PM)		

Statement on Academic Conduct and Support Systems

Academic Conduct:

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” policy.usc.edu/scampus-part-b. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the Office of Equity and Diversity <http://equity.usc.edu> or to the Department of Public Safety <http://capsnet.usc.edu/department/departement-public-safety/online-forms/contact-us>. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. The Center for Women and Men <http://www.usc.edu/student-affairs/cwm/> provides 24/7 confidential support, and the sexual assault resource center webpage <http://sarc.usc.edu> describes reporting options and other resources.

Support Systems:

Student Health Counseling Services - (213) 740-7711 – 24/7 on call
engemannshc.usc.edu/counseling

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call
suicidepreventionlifeline.org

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-4900 – 24/7 on call
engemannshc.usc.edu/rsvp

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED) | Title IX - (213) 740-5086
equity.usc.edu, titleix.usc.edu

Information about how to get help or help a survivor of harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations.

Bias Assessment Response and Support - (213) 740-2421
studentaffairs.usc.edu/bias-assessment-response-support

Avenue to report incidents of bias, hate crimes, and microaggressions for appropriate investigation and response.

The Office of Disability Services and Programs - (213) 740-0776

dsp.usc.edu

Support and accommodations for students with disabilities. Services include assistance in providing readers/notetakers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710

studentaffairs.usc.edu/ssa

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

diversity.usc.edu

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

dps.usc.edu, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call

dps.usc.edu

Non-emergency assistance or information.