

**ISE 535 Data Mining**

Spring 2020

**Location:** Online

**Instructor:** Bruce Wilcox

**Office:** GER 216

**Office Hours:** Tuesdays, 7:00-8:00PM

**Contact Info:** [brucewil@usc.edu](mailto:brucewil@usc.edu)

**Teaching Assistant:** Vinith Angadi

**Office:** N/A

**Office Hours:** Thursdays, 12:00 – 2:00PM

**Contact Info:** [vangadi@usc.edu](mailto:vangadi@usc.edu)

**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 both the preprocessing and visualization 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.
- Explore the data mining disciplines using business-oriented case studies with “mini-projects” that provide the students with experience that will assist in their post-graduation activities to secure and perform in jobs in the data mining field.
- *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**

There are no required texts. The lectures and associated PowerPoints, in-class exercises, and projects are intended to be complete summaries of class content. However, students are strongly encouraged to consult the following texts for supplementary materials as referenced in the syllabus and in class:

- Tan, et. al., *Introduction to Data Mining*, Pearson Education, 2019 (IDM)
- Torgo L., *Data Mining with R*, CRC Press, 2017 (DMR)
- James, et. al., *An Introduction to Statistical Learning with Applications in R*, Springer, 2017 (ISL)
- 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

- **R assignment** (to ensure that you have achieved basic proficiency with this tool) – 10% of final grade
- **Five homework projects (one per module for modules 2-6)** - 50% of final grade
- **One final course project** that will tie together several of the concepts presented in the class. - 40% of final grade

## Grading Policy

Assignment	% of Grade
R assignment (1)	10
Homework Projects (5)	50
Final Exam/Project	40
<b>TOTAL</b>	<b>100</b>

**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

## Assignment Submission Policy

Assignments should be typewritten and clean. 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	Homework	Reference
1	Aug 28	Module 1: Introduction Introduction to R, RStudio, and R Markdown.		
2	Sep 4	Module 2: Data Exploration and Preparation		IDM Ch. 2
3	Sep 11	Module 2: Data Exploration and Preparation	R Assignment Due	IDM Ch. 2
4	Sep 18	Module 2: Data Exploration and Preparation		IDM Ch. 2
5	Sep 11	Module 3: Classification	Module 2 HW Due	ISLR Ch. 8 IDM Ch. 3-4
6	Sep 18	Module 3: Classification		ISLR Ch. 8 IDM Ch. 3-4
7	Oct 2	Module 4: Clustering	Module 3 HW Due	ISLR Ch. 10 IDM Ch. 7-8
8	Oct 9	Module 4: Clustering		ISLR Ch. 10 IDM Ch. 7-8
9	Oct 16	Module 5: Association Rule Analysis	Module 4 HW Due	IDM Ch. 5 IDM Ch. 5-6
10	Oct 25	Module 5: Association Rule Analysis		IDM Ch. 5 IDM Ch 5-6
11	Nov 6	Module 6: Anomaly Detection	Module 5 HW Due	IDM Ch. 9
12	Nov 13	Module 6: Anomaly Detection		IDM Ch. 9
13	Nov 20	Class review and final project work	Module 6 HW Due	
14	Dec 4	Final Exam/Project Due	Final Project Due	

## 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](http://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](http://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/departments/departments-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](http://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](http://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](http://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](http://equity.usc.edu), [titleix.usc.edu](http://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](http://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](http://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/ssu](http://studentaffairs.usc.edu/ssu)

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](http://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](http://dps.usc.edu), [emergency.usc.edu](http://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](http://dps.usc.edu)

Non-emergency assistance or information.