USC Viterbi
School of Engineering

ISE 535: Data Mining
4 Units

Day/Time: Mon, Wed 10:00-11:50am, Spring 2025
Location: KAP 148

Instructor: Dr. Tao Ma
Office: GER 216
Office Hours: TBD or by appointment via email

Contact Info: ma.tao@usc.edu

Teaching Assistant: TBD

Catalog Description

Data preprocessing, cleaning and visualization, statistical inference and testing, data classification, clustering analysis, association pattern mining, data streams mining, anomaly detection and false discoveries.

Course Description

Data mining is the discipline of extracting useful insights from large quantities of data. As such, the focus in this class is on statistical inference but not on prediction (which is the topic of ISE-529).

This course is organized into two broad sections:

- Exploratory data analysis and statistical data analysis techniques to find useful information from data.
- Algorithm-based data mining techniques, statistical (machine) learning-based techniques for data classification, clustering analysis, association pattern mining, data streams mining, outlier analysis.

To the maximum extent possible, this course teaches the concepts by means of case studies with real-world data and examples.

Learning Objectives and Outcomes

- Develop an advanced level of proficiency with the preprocessing, visualization, and statistical analysis of data as well as the primary data mining algorithmic techniques.
- Review and re-enforce basic statistical concepts that are important in the field of data science.
- Apply unsupervised modeling techniques (including clustering and association rule mining) to analyze and obtain insights from data.
- Take raw data and perform all of the steps necessary to generate a professional exploratory data analysis report.

Class Delivery Mode: This class will be delivered via classroom lectures, homework assignments, exams and/or project. Students are required to attend lectures in person. Exams must be taken in person.

Prerequisite(s): None

Recommended Preparation:
Undergraduate course in statistics and working knowledge of a programming language

Course Policies

All course materials (lecture slides, homework, exercises, etc.) will be distributed via Brightspace. They are only for the students who are enrolled in the course. Do **NOT** post or duplicate the course materials online or distribute them electronically or in print without prior explicit permission from the instructor. All assignments will be submitted through Brightspace.

Technological Proficiency and Hardware/Software Required

This course will utilize the Python programming language and associated libraries which are open source and available to the students for no cost.

Textbooks

The theoretical materials in the course are drawn from the following texts:

- Pang-Ning Tan, et. al., Introduction to Data Mining 2nd ed. 2019 ISBN 978-0-13-312890-1
- Charu C. Aggarwal, Data Mining, Springer, 2015 ISBN 9783319141411
- Bruce, et. al., Practical Statistics for Data Scientists, O'Reilly, 2020 (PSDS)

Grading Breakdown

The course grade distribution is as follows:

- Homework assignments (approximately 6-8) – 50% of final grade
- Midterm exam (in class) – 20% of final grade
- Final exam or project – 30% of final grade

The mid-term and final exam will be held in-person during the class time. They will be closed book with one-page of cheat sheet permitted. You will be allowed to choose between final exam and project. Details on the final project will be released the week of the mid-term.

Grading Scale

Course final grades will be determined using the following scale

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

Borderline averages between two letter grades may be rounded up based on class engagement at the instructor's discretion.

Assignment Submission Policy, Timelines, and Rules for Submission

- Assignments will be posted on Brightspace and submitted to Brightspace by the due date for grading.
- Late submissions are NOT accepted after the due date or will incur a 15% penalty for special cases.
- No submissions will be accepted after 48 hours past the due date and assignments not submitted will result in 0 grade.
- The lowest homework grade for the semester will be dropped from the final grade computation.
- No make-up exams are considered. If missing exams, you will receive 0 grade.

Course Schedule

The following table is a weekly breakdown tentative schedule and subject to change according to the actual class situation throughout the semester. Please follow the announcement in class or Brightspace for the latest update.

- Module 1 - Introduction to Data Mining (1 week)
  - Data/feature types
  - Data structuring and cleansing
- Module 2 - Foundations of Exploratory Data Analysis (2 weeks)
  - Introduction to EDA and descriptive statistics
  - Data visualization techniques
  - Handling missing and anomalous data.
  - Generating comprehensive EDA reports
  - Feature distributions and relationships
- Module 3 - Statistical Foundations for Analytics (2 weeks)
  - Statistical inference and hypothesis testing
  - Computational statistics (resampling, bootstrapping, permutation testing)
  - Probability distributions for analytics
  - Statistical assumptions and model validity
  - Avoiding false discoveries and understanding p-hacking
- Module 4 - Inference and Interpretability in Supervised Learning (2-3 weeks)
  - Overview of model transparency
  - Global interpretation techniques
  - Local interpretation techniques
  - Model-agnostic vs model-specific transparency techniques.
- Module 5 - Finding Structure in Data: Unsupervised Learning Methods (2-3 weeks)
  - Clustering techniques
  - Dimensionality reduction
  - Anomaly detection
- Module 6 - Pattern Discovery with Association Rules (2 weeks)
  - Key metrics for association rules
  - Algorithms for mining association rules
  - Applications and use cases.
- Module 7 - Mining Data Streams (Time permitting)

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/department-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.