USCViterbi
School of Engineering

ISE 535 Data Mining
Units: 3
Spring 2021- Wednesday 4 – 7 p.m.
Location: TBD

Instructor: Cesar Acosta
Office: GER 216
Office Hours: TBD
Contact Info: acostame@usc.edu

Teaching Assistant: TBA
Office: TBA
Office Hours: TBA
Contact Info: TBA

IT Help:
Hours of Service:
Contact Info:

Catalog description

Data preprocessing, data cleaning, data summarization, data visualization, and predictive modeling for classification and regression. Modeling dependencies using association rules.

Course Description

This course is about data analytics tools, methods, and applications. It focuses on data preprocessing, data wrangling, and data cleaning of dataframes and text data. Data Mining on the web is also included. The course shows how to do feature engineering and how to reduce data complexity. Regression methods include KNN and SVM for regression.

Data visualization techniques are reviewed to find useful information from spatial data, now available from different on line providers.

Unsupervised Learning Methods are used for Clustering analysis. The course reviews several clustering methods such as K-Means and Hierarchical clustering.

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

Learning Objectives and Outcomes

In this course students learn to

- Preprocess dataframes (missing, duplicates, and data types) and text files
- Perform simple and advanced A/B testing to compare populations means and populations proportions.
- Build Time Series models for forecasting.
- Apply clustering methods for unsupervised learning.
- Perform Data Mining on the Web
- Apply data visualization tools for descriptive and predictive analytics on spatial data.
- Use association rules for data mining and modeling.

Prerequisite(s): None.

Recommended Preparation Expected to have knowledge of Engineering Statistics on the level of ISE 225 and working knowledge of a programming language.

Course Notes
The course material is available on Blackboard.

Technological Proficiency and Hardware/Software Required
Required software: The R programming language will be used.

Required Textbook
None

Supplementary Materials (References)

- James G., *An Introduction to Statistical Learning*, Springer, 2013 (ISLR)
  ISBN 978-1-4614-7137-0
- Torgo L., *Data Mining with R*, CRC Press, 2017, ISBN 978-1-4822-3489-3 (DMR)
- Kuhn M., Johnson K., *Applied Predictive Modeling*, Springer, 2013,
  ISBN 978-1-4614-6849-3 (APM)
- Chapman C., McDonnell Feit E., *R for Marketing Research and Analytics,* Springer 2015, ISBN 978-3-319-14436-8 (RMRA), available from the Science library as an e-book.

Description and Assessment of Assignments

- **Midterm** will be in-class based on the schedule and 2 hours length.
- **Final Examination** a two-hour comprehensive exam scheduled by USC.
- **Homework** are assigned every other week. Homework is based on the material of the previous and current week. Must be submitted by the due date, during the class session. No late homework to be accepted.

Grading Policy

Assignment	Points	% of Grade
Homework	100 each (6 homework assignments)	30
Midterm	100	30
Final	100	40
TOTAL		100

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 in class by the due date. Email submissions and late submissions are not allowed. No make-up exams are considered.

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	Topics/Daily Activities	Homework	Reference
1	**Introduction to Data Mining** for Descriptive and Predictive Analytics. **Introduction to R,** RStudio**,** and rmarkdown.		DMR Ch 2, Ch3.6.
2	**Data Preprocessing.** Data cleaning, Tidy data, formats, dates, NA values.	HW2 due	DMR Ch3.
3	**Time Series Analysis.** Autocorrelation, Partial autocorrelation, Time Series decomposition, Time Series cross validation, AR, AM, ARMA models.	HW3	DMR Ch3.
4	**Regression topics.** KNN for regression. SVM for regression.	HW3 due	DMR Ch3, Ch6.
5	**Classification.** Discriminant Analysis and the multivariate normal distribution. Naive Bayes. Classification trees, random forests, and boosting with R.		ISLR Ch 4.
6	**A|B Testing.** Part 1. Comparing two populations. The two-sample t test. Matched Pairs Experiment (the paired t test). Difference between two means and the difference between two population proportions		Notes.
7	**A|B Testing.** Part 2. Difference among many designs. The multi-arm bandit algorithm. Advanced A|B testing.		Notes.
8	**Midterm Exam**		
9	**Unsupervised Learning.** Principal Component Analysis. Principal Components Regression.		ISLR Ch 10.
10	**Unsupervised Learning.** Clustering Methods. K-Means clustering. Hierarchical clustering.		ISLR Ch 10.
11	**Data Mining on the Web.** Web Scrapping with R. Google Analytics Data with R.		**Notes.**
12	**Data Visualization.** R library ggplot2.		**Notes.**
13	**Data Visualization.** R library ggmap. Spatial and geographical visualization. Google Map Static API. R library rworldmap.	HW6	RMRA Ch 11.
14	**Bayesian Computation with R.**		Notes.
15	**Data Modeling**. Association rules for Market Basket Analysis.	HW6 due	RMRA Ch 12.
16	**Final Exam**		

Date and time of the final for this class shown in the USC *Schedule of Classes* at classes.usc.edu/.

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.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
studenthealth.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 and Services (RSVP) - (213) 740-9355(WELL), press "0" after hours – 24/7 on call
studenthealth.usc.edu/sexual-assault
Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298
equity.usc.edu, titleix.usc.edu
Information about how to get help or help someone affected by 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. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
usc-advocate.symplicity.com/care_report
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity |Title IX for appropriate investigation, supportive measures, 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
uscsa.usc.edu
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.