



SIMON
BUSINESS SCHOOL
UNIVERSITY of ROCHESTER

CIS 434 LAB: SOCIAL MEDIA AND TEXT ANALYTICS LAB

Spring B 2022 (March 23 – May 4th)

Instructor: Junyuan Ke	Times: Wed 13:30–16:20
Email: junyuan.ke@simon.rochester.edu	Room: TBA

Course Pages:

1. [Course Blackboard Link](#) (The Zoom link for class will be found there. Lectures slides will be uploaded before each class).

Office Hours: TBD; After class, or by appointment.

Main References: The lecture notes distributed via Blackboard are self-contained and hence there is **no required textbook for this course**. The Blackboard course site will serve as a repository for class materials. Current schedule and reading list, homework and lecture notes will all be available from it. All Zoom sessions will be recorded and made available on Blackboard. Simon IT will provide all students registered for the course with credentials to access the site.

The following books are helpful but not required for the class:

- [Murphy, Kevin. Machine Learning: A Probabilistic Perspective. 2012.](#)
- [Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly, 2017](#)
- [Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. Prediction Machines: The Simple Economics of Artificial Intelligence. Harvard Business Press, 2018.](#)

Course Description and Learning Objectives: This is a complementary lab to the CIS434 main course. This lab reinforces the mathematical and statistical knowledge MS/MBA students learned in the main course and provides sufficient hands on coding experience in manipulating unstructured data and properly utilizing prediction models.

Relation with other Courses at Simon Business School: The objective of this course is to develop and sharpen a quantitative foundation and coding ability that students can use to complete the Simon MBA/MS program. The course material will start from basic python coding and gradually cover all the necessary mathematical and statistical concepts and if time permits, will cover more advanced topics in Machine Learning. Experience in Python/R coding and Machine Learning is not required but strongly recommended, I recommend taking at least one course among [GBA465, GBA468, CIS432], before or at least in parallel with this course.

Grading Policy*: * We will follow grading policy of the main course.

Class Policy: Regular attendance is essential and expected.

Academic Integrity: Simon’s Code of Academic Integrity (see Section 2 of the Student Handbook) states: “Every Simon student is expected to be completely honest in all academic matters. Simon students will not in any way misrepresent their academic work or attempt to advance their academic position through fraudulent or unauthorized means. No Simon student will be involved knowingly with another student’s violation of this standard of honest behavior.” Issues pertaining to academic integrity will be dealt with according to the Simon school code on academic integrity.

Tentative Course Outline:

Topic 1: Python Basics

Handling text data with Pandas and Numpy, text cleaning techniques, word frequency generation, and lexicon-based sentiment analysis packages.

Topic 2: Predictive Models and Overfitting

Fundamental Machine Learning pipeline, train-test splitting techniques, predictive models such as logistic regression, decision trees, naive Bayes classifier, and SVM, different performance metrics such as precision recall, and the idea of bias-variance trade off. We will also go over overfitting in a separate example and introduce k-fold and LOO cross validation techniques.

Topic 3: LDA Practice and Interpretation

Perform Topic Modeling with LDA and techniques to finding interpretable insights from Topic Modeling. We will also go over tips students can use to get better results in assignment 2.

Topic 4: Word2Vec Practice Similarity Understanding and implementation of W2V. Tips on generating business insights with word similarity. We will also go over tips students can use to get better results in assignment 3.

Topic 5: ML Practice

More advanced ML models. More tips and Q&A on assignment 3.

Topic 5: NN Practice

Hands on practice on deep learning models. We will also go over tips students can use to get better results in assignment 1.

Topic 6: NN Practice and Group Presentation

More advanced topics such as interpretable ML if we have time. Q&A session on group project before students’ presentation.