

Introduction to Text Analysis

Summer Institute

July-August 2025

Time: July 28 - August 7, 2025
Live classes Monday, Tuesday, and Thursday 10:00-11:30 (Eastern Time)

Location: Zoom

Instructor: Robyn Ferg, Ph.D.
Westat
fergr@umich.edu

Overview:

In this two-week course, students will learn a variety of natural language processing methods for analyzing and extracting meaning from text data. The course will start with an introduction to text data, including text preprocessing and exploratory methods. The topics that follow will include machine learning models used for topic modeling, clustering, classification, sentiment analysis, and word embeddings. Students will also be introduced to web scraping. Considerations to both long and short texts of various subject matter. Class examples will be demonstrated primarily in R. This course assumes a bachelors-level background in Statistics or related field and knowledge of R; no prior knowledge of text analysis is assumed.

Course Materials:

All course materials will be available online on the class Canvas website. Course materials on the website include video lectures, recommended readings, assignments, and code.

Accommodations for Students with Disabilities:

If you believe you need an accommodation for a disability, please contact the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; <http://ssd.umich.edu>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and will remain confidential.

Course Schedule

Monday, July 28

10:00-11:30 Course introduction, student introductions, course expectations

Lecture: Introduction to Text Data

Introduction to text data code examples

Assignment 1 released

Tuesday, July 29

10:00-11:30 Lecture: Topic Modeling

Topic modeling code examples

Wednesday, July 30

Video Lecture:

Lecture: Topic Modeling Extensions

Code for topic modeling extensions

Lecture: Dimension Reduction

Code for dimension reduction

Thursday, July 31

10:00-11:30 Lecture: Word Embeddings

Code for word embeddings

Assignment 2 released

Friday, August 1

Assignment 1 due

Monday, August 4

10:00-11:30 Classification

Classification code

Assignment 2 due

Tuesday, August 5

10:00-11:30 Clustering

Clustering code

Assignment 3 released

Wednesday, August 6

Video Lecture: Web Scraping

Web scraping code
API code
Course wrap-up

Thursday, August 7

10:00-11:30 Lecture: Sentiment Analysis
Sentiment analysis code
Lecture: Text Summarization
Text summarization code

Assignment 3 due