

PL/CS 362-1: Applied Computational Methods for Social Sciences

Details

Instructor: Bogdan G. Popescu

Hours: MW 10:00-11:15AM

Total Hours of Contact: 2:30 per week

Room: Mimose 3

Credits: 3

Prerequisites: None

Office Hours: TBA

Course Description

This is an introductory course on using Python, a free and versatile programming language, to analyze text data. The course begins by building foundational programming skills in Python, including working with variables, loops, conditional statements, and data structures like lists, dictionaries, and arrays. In the second part, students will explore Python's powerful tools for analyzing text data, such as social media posts, email correspondence, customer reviews, and political debates. The course also includes an introduction to ChatGPT and prompt engineering, teaching students how to leverage large language models for advanced text analysis tasks.

Through hands-on exercises, lectures, and projects, students will learn to process, clean, and analyze textual data systematically and reproducibly. By the end of the course, participants will be equipped with the skills to derive meaningful insights from textual information and integrate generative AI tools like ChatGPT into their workflows.

Note: You should have a computer either Windows or Mac with at least 8GB of RAM memory.

Summary of Course Content

This course provides a comprehensive introduction to Python programming and its applications in text analysis. The initial portion of the course covers the fundamentals of Python, including working with variables, data structures, and basic programming constructs. The latter part focuses on Python's capabilities in text data analysis, including text cleaning, dictionary methods, sentiment analysis, topic modeling, and machine learning-based approaches.

Additionally, the course introduces ChatGPT and prompt engineering, equipping students to use large language models for tasks such as text summarization, classification, sentiment analysis, and

topic modeling. Tutorials on creating professional portfolios and websites further enhance students' ability to showcase their analytical skills. The knowledge and techniques learned in this course are applicable across diverse fields, including political science, economics, business, and marketing.

Learning Outcomes

Upon successful completion of this course the students will be able to:

- Write Python programs to perform tasks such as loops, conditional statements, and function definitions.
- Employ quantitative techniques to process and analyze textual data.
- Utilize Python libraries like Numpy, Pandas, and NLTK for text manipulation and analysis.
- Apply advanced methods such as sentiment analysis, topic modeling, word embeddings, and supervised learning to text data.
- Use ChatGPT and prompt engineering to enhance text analysis tasks, including summarization, classification, and generating structured outputs.
- Create and publish a professional website showcasing their portfolio and analytical capabilities.

Assessment

Assessment methods:

You will be graded on four problem sets during the semester (each 12.5% of your grade) and a final report and presentation (each 25% of your grade).

- 4 problem sets: 50% of the final grade (12.5% each)
- final presentation: 25% of the final grade
- final project report: 25% of the final grade

Problem Sets

1.Initial Individual Submission: This component contributes 50% of the overall grade for the problem set. When you complete the problem set independently and submit it to the instructor, your grade for this component will be calculated based on the quality of your independent work. This grade will be weighted at 50% of the total assignment grade.

2.Final Submission After Group Consultation: This component also contributes 50% of the overall grade for the problem sets. After discussing the problem set with your group members and documenting the correct answers, you will submit this revised version individually. Note: no group submission is permitted. Each one of you has to submit the second attempt of the assignment individually. Your grade for this component will be based on the quality of your final submission after group consultation. This grade will also be weighted at 50% of the total assignment grade.

Presentation

Each student will deliver a 10-15 minute presentation on a topic assigned in advance. Presentations should include an overview of the:

- question
- description of the selected text analysis method
- description of the dataset used
- description of how the data was validated
- critical analysis

Please deliver your presentations in a quarto presentation and use relevant visuals. Practice beforehand to stay within the time limit and maintain a confident, professional tone. Be prepared to answer 2-3 questions from peers or the instructor during and after the presentation. Remember to cite your sources and avoid reading verbatim from slides or notes.

In addition to summarizing the key arguments or findings, your presentation should include critical analysis of the material. Highlight what the author does not address, the limitations of their research, or potential problems in their analysis or methodology. Think about how the research could be improved, expanded, or connected to broader themes discussed in class, and incorporate these insights into your presentation.

Final Project

Students will undertake a text analysis project emphasizing the practical application of text analysis techniques using the python programming language. This project offers an opportunity to showcase the acquired skills in manipulating text data and conducting meaningful analyses. Participants are encouraged to choose a topic of interest or relevance, utilizing datasets provided in the course or exploring new ones. The project entails a few steps:

- choosing a topic that involves text data
- acquire data either from the course materials or from external sources

Students have to submit a two-page report in Quarto and give a 20-minute and give an in-class presentation in Quarto.

Grading Criteria for the project:

- Relevance
- Methodology
- Analysis
- Presentation
- Code Quality

Here are examples of projects that previous students created:

- [Topical Similarity in the US Presidential Speeches](#)
- [Framing Latin America in UN Speeches](#)
- [Agreesion in Political Speeches](#).

Attendance Requirements

Students are required to attend classes following the University's policies. Students with more than two unexcused absences are assumed to have withdrawn from the course. Students with a justified reason not to attend class have to send me an email explaining why they cannot attend ahead of class and need to submit a [form to the Dean's Office](#).

Academic Honesty

As stated in the university catalog, any student who commits an act of academic dishonesty will receive a failing grade on the work in which the dishonesty occurred. In addition, acts of academic dishonesty, irrespective of the weight of the assignment, may result in the student receiving a failing grade in the course. Instances of academic dishonesty will be reported to the Dean of Academic Affairs. A student who is reported twice for academic dishonesty is subject to summary dismissal from the University. In such a case, the Academic Council will then make a recommendation to the President, who will make the final decision.

Students with Learning or Other Disabilities

John Cabot University does not discriminate on the basis of disability or handicap. Students with approved accommodations must inform their professors at the beginning of the term. Please see the website for the complete policy

Required Books

General Books on Python

McKinney, Wes, 2022. *Python for Data Analysis*. O'Reilly. <https://wesmckinney.com/book/>.

Text Analysis

Hovy, Dirk. 2020. *Text Analysis in Python for Social Scientists*. Cambridge University Press. <https://www.cambridge.org/core/elements/abs/text-analysis-in-python-for-social-scientists/BFAB0A3604C7E29F6198EA2F7941DFF3>

Bengfort, Benjamin, Bilbro, Rebecca, and Tony Ojeda. 2018. *Applied Text Analysis with Python*. O'Reilly Media, Inc. <https://www.oreilly.com/library/view/applied-text-analysis/9781491963036/>

Hvitfeldt, Emil and Silge, Julia. 2022. *Supervised Machine Learning for Text Analysis in R*. CRC Press. <https://smltar.com>

Grimmer, Justin, Brandon M. Stewart, and Margaret E. Roberts. *Text As Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press, 2022.

Silge, Julia, and David Robinson. *Text Mining with R: A Tidy Approach..* First edition. O'Reilly, 2017.

Week 1

Class 1: Intro to Python and Text Analysis

01/20/2025 - Mon - [Lecture](#)

- Introduction and course overview
- Installing R, R Studio, and Anaconda
- Introduction to Python programming and its applications in text analysis
- Overview of text data (e.g., reviews, political debates, translations)

Class 2: Intro to Python, Jupyter notebooks, and R Quarto

01/22/2025 - Wed - [Lecture](#)

- Quarto files: notebooks vs. presentations
- The Python environment
- [How to Make Slides in Quarto](#)

Reading

McKinney, Wes, 2022. *Python for Data Analysis*. O'Reilly. <https://wesmckinney.com/book/>. C1-2

Week 2

Class 1: Variables, Strings, Numbers

01/27/2025 - Mon - [Lecture](#)

- Variables
- Strings
- Combining strings
- Numbers
- Booleans
- Conditionals: `if` and `elif`

Class 2: Loops, Lists, Breaks, and Zip Functions

01/29/2025 - Wed - [Lecture](#)

- Comparison Operators
- Error Handling
- Lists
- Loops
- Breaks

Reading

McKinney, Wes, 2022. *Python for Data Analysis*. O'Reilly. <https://wesmckinney.com/book/>. C3

[Assignment 1](#)

Week 3

Class 1: Operations with Lists and Tuples

02/03/2025 - Mon - [Lecture](#)

- Lists
- List Operations
- List Slicing
- Tuples

Class 2: While Loops and Functions

02/05/2025 - Wed - [Lecture](#)

- While loops
- Functions

Reading

McKinney, Wes, 2022. *Python for Data Analysis*. O'Reilly. <https://wesmckinney.com/book/>. C3

Week 4

Class 1: Dictionaries and Sets

02/10/2025 - Mon - [Lecture](#)

- dictionaries
- extracting elements from dictionary
- sorting dictionary keys
- nesting dictionaries
- dictionaries within dictionaries
- sets and operations with sets
- Supplementary Tutorial: [Intro to Numpy](#)

Class 2: Pandas

02/12/2025 - Wed - [Lecture](#)

- Dataframes
- Dataframe Methods
- Missing Data
- Assigning Values

Week 5

Class 1: Pandas Data Wrangling

02/17/2025 - Mon - [Lecture](#)

- Setting Indexes
- Changing, ordering column names
- Identifying and Removing duplicates
- Filtering Data
- Aggregating Data
- Modifying Data Structures

Class 2: Merging Data

02/19/2025 - Wed - [Lecture](#)

- Aggregating by Group
- Inner, Outer Joins
- Left, Right Joins

[Assignment 2](#)

Week 6

Class 1: Data Visualization 1

02/24/2025 - Mon - [Lecture](#)

- ggplot Basics
- Simple Geoms vs. Complex Geoms
- Layers

Class 2: Data Visualization 2

02/26/2025 - Wed - [Lecture 1](#) | [Lecture 2](#)

- Colors, Shapes, Transparency
- Facets
- Coordinates
- Annotations

Week 7

Class 1: Text Analysis Intro

03/03/2025 - Mon - [Lecture](#)

- Latent Concepts
- Dictionaries, Supervised learning, text scaling, topic models, word embeddings
- Workflows of Text Analysis
- Documents, features,
- Bag-of-Words and Document-feature matrix
- Text Cleaning in Python

Class 2: Intro to ChatGPT for Text Analysis

03/05/2025 - Wed - [Lecture](#)

- Prompt engineering
- Personas, delimiters, structured outputs, Programmatic consumptions
- Chain of Thought
- ChatGPT Api, pricing

Week 8

Class 1: Applied ChatGPT for Text Analysis

03/17/2025 - Mon - [Lecture](#)

- Text Summarization, Text Classification
- Sentiment Analysis, Emotion Analysis
- Topic Modeling, Topics using Zero-Shot Learning
- Language, Grammar and Spelling

Class 2: Dictionary Methods

03/19/2025 - Wed - [Lecture](#)

- Counting words
- Examples of Dictionaries
- Validation of Dictionary Methods
- Accuracy, Sensitivity, Specificity, Naive Guess

Week 9

Class 1: Similarity

03/24/2025 - Mon - [Lecture](#)

- Inner Product
- Frequency words, Euclidian distance
- Cosine Similarity

Class 2: Similarity and Word Clouds

03/26/2025 - Wed - [Lecture](#)

- Similarity: Levenshtein distance
- Cosine similarity
- TF-IDF, Bag-of-Words
- Word Clouds, Fightin' Words

[Assignment 3](#)

Week 10

Class 1a: Designing Projects using Text Analysis

03/31/2025 - Mon - [Lecture](#)

Class 1b: Language Complexity

[Lecture](#)

- Flesch Reading Ease
- Kincaid score

Class 2: Supervised Learning

04/02/2025 - Wed - [Lecture](#)

- Naive Bayes
- Language Models
- Validating Supervised Learning, K-Fold Validation

Week 11

Class 1: Unsupervised Learning: Topic Models

04/07/2025 - Mon - [Lecture](#)

- Topic Models
- Latent Dirichlet Allocation
- Semantic Coherence

Class 2: Unsupervised Learning: Word Embeddings

04/09/2025 - Wed - [Lecture](#)

- Word Embeddings
- Distributional Semantics, Topic Distributions
- BERTopic

[Assignment 4](#)

Week 12

Class 1: Student presentations

04/14/2025 - Mon

- Blumenau, Jack E; Hargrave, Lotte, 2022. [No Longer Conforming to Stereotypes? Gender, Political Style, and Parliamentary Debate in the UK](#). British Journal of Political Science - **Wendy and Nico**
- Blaydes Lisa; Grimmer, Justin; McQueen, Alison, 2018. [Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds](#). Journal of Politics - **Harman**
- Bellodi, Luca, 2022. [A Dynamic Measure of Bureaucratic Reputation: New Data for New Theory](#). American Journal of Political Science - **Kayla**
- Motolinia, Lucia, 2020. [Electoral Accountability and Particularistic Legislation: Evidence from an Electoral Reform in Mexico](#). American Journal of Political Science - **Catalina**
- Riccardo Di Leo, Chen Zeng, Elias Dinas, and Reda Tamtam, 2025. [Mapping \(A\)Ideology: A Taxonomy of European Parties Using Generative LLMs as Zero-Shot Learners](#). Political Analysis - **Martina**

Class 2: Student Proposal Presentations

04/16/2025 - Wed

1. Question
2. Data sources
3. Proposed Text Analysis functions
4. Expected policy implications

Week 13

Class 1: Sentiment Analysis

04/21/2025 - Mon - [Lecture](#)

- Lexicon-based approaches, machine learning
- VADER sentiment analysis, transformers
- Zero-shot classification

Class 2: Making a Quarto Website

04/23/2025 - Wed - [Lecture](#)

- Making a website on Quarto
- Storing a website on GitHub
- Structuring a professional website for Data Analytics Jobs
- Website Troubleshooting

Week 14

Class 1: Course Reflection

04/28/2025 - Mon - [Lecture](#)

- Overview of the Course
- Skills Acquired
- Jobs where these skills are valued

Class 2: Project Development Time

04/30/2025 - Wed