## Course Information

**Course Title:** Computational Text Analytics in Python
**Course Number:** JOUR 779P/479P
**Term:** Fall 2023
**Credits:** 3
**Class Time:** 9:30 AM - 12:15 PM
**Location:** KNI 1101

**Professor:** Dr. Naeemul Hassan
**Email:** nhassan@umd.edu
**Office Hours:** Thu 4:00 PM - 5:00 PM
**Office Zoom Link:**
https://umd.zoom.us/j/3503870486?pwd=bkdHSVNTb1kxVkIyK2tHRVZ2R25sUT09

## Course Description

Processing text data is crucial in many domains such as computer science, journalism, social science, psychology, political science, etc. In this era of the internet and social media, data is generated in such a huge volume and at such a high speed that it is practically impossible to process it and find insightful patterns from it in traditional ways. That is why students, instructors, and researchers of various domains are embracing computational tools to perform statistical textual content analysis. This course will enable you to perform large-scale quantitative analysis of textual data in an authoritative way and find useful patterns from the data. It is designed to teach the principles and computational methods of natural language or textual data processing and provide hands-on experience in text analysis using Python. Natural language processing (NLP) is a set of techniques for examining and analyzing text as data. This course investigates the usage of NLP as a set of these techniques, with a particular emphasis on the practical side of NLP—using current NLP methods and libraries in Python in novel and inventive ways. The representation of text, including features derived from linguistic structure (such as parts of speech, named entities, syntax, and coreference), and features derived from low-dimensional representations of words, sentences, and documents are some of the topics covered. Other topics include exploring textual similarity and information extraction. This course will concentrate on both traditional methods such as Tf-iDf as well as more recent techniques such as BERT.

## Learning Outcomes

After successfully completing this course you will be able to:

+ Describe the fundamental concepts and techniques of natural language processing or text processing using Python.
+ Distinguish among the various text processing techniques, taking into account the assumptions, strengths, and weaknesses of each.
+ Analyze real-world problems and model them for the application of statistical text processing techniques.
+ Use appropriate descriptions, visualizations, and statistics to communicate problems and solutions.
+ Analyze large volumes of text data generated from a range of real-world applications using Python.
+ Understand the ethical implications of a computing solution or algorithm. Understand the human factors that should be considered in designing a text analytics algorithm.

## Reference Books and Resources

- Natural Language Processing with Python– Analyzing Text with the Natural Language Toolkit
  Steven Bird, Ewan Klein, and Edward Loper
  Free Copy: https://www.nltk.org/book/

- Speech and Language Processing (3rd ed. draft)
  Dan Jurafsky and James H. Martin
  Free Copy: https://web.stanford.edu/~jurafsky/slp3/

Other resources will be provided in the class by the instructor

## Course Structure

- In person. Location KNI 1101

## Tips for Success in an Online Course

1. **Participate.** Discussions and group work are critical parts of the course. You can learn a great deal from discussing ideas and perspectives with your peers and professor. Participation can also help you articulate your thoughts and develop critical thinking skills.
2. **Manage your time.** Make time for your online learning and participation in discussions each week. Give yourself plenty of time to complete assignments including extra time to handle any technology-related problems.
3. **Login regularly.** Log in to ELMS-Canvas several times a week to view announcements, discussion posts, and replies to your posts. You may need to log in multiple times a day when group submissions are due.
4. **Do not fall behind.** This class moves at a quick pace and each week builds on the previous. It will be hard to keep up with the course content if you fall behind in the pre-work or post-work.
5. **Use ELMS-Canvas notification settings.** Canvas ELMS-Canvas can ensure you receive timely notifications in your email or via text. Be sure to enable announcements to be sent instantly or daily.
6. **Ask for help if needed.** If you need help with ELMS-Canvas or other technology, IT Support. If you are struggling with a course concept, reach out to me, and your classmates, for support.

## Campus Policies

It is our shared responsibility to know and abide by the University of Maryland's policies that relate to all courses, which include topics like

- Academic integrity
- Student and instructor conduct
- Accessibility and accommodations
- Attendance and excused absences
- Grades and appeals
- Copyright and intellectual property

Please visit www.ugst.umd.edu/courserelatedpolicies.html for the Office of Undergraduate Studies' full list of campus-wide policies and follow up with me if you have questions.

## Course Guidelines

### Academic Integrity

Note to you the instructor: In this course, we will collect written assignments using Turnitin's "Originality Checker." This tool scans student submission against online content and previously-submitted papers, alerts students to some writing errors (e.g., incorrect or insufficient citations) so that they can improve their writing, and alerts faculty to submissions that may contain text matching another source. For

guidance on academic integrity and how to use Turnitin in your course, visit the TLTC's page on Academic Integrity & Technology. If you decide to use Turnitin, below is some sample syllabus language you can include:

For this course, some of your assignments will be collected via Turnitin on our course ELMS page. I have chosen to use this tool because it can help you improve your scholarly writing and help me verify the integrity of student work. For information about Turnitin, how it works, and the feedback reports you may have access to, visit Turnitin Originality Checker for Students

## Names/Pronouns and Self-Identifications

The University of Maryland recognizes the importance of a diverse student body, and we are committed to fostering inclusive and equitable classroom environments. I invite you, if you wish, to tell us how you want to be referred to both in terms of your name and your pronouns (he/him, she/her, they/them, etc.). The pronouns someone indicates are not necessarily indicative of their gender identity. Visit trans.umd.edu to learn more.

Additionally, how you identify in terms of your gender, race, class, sexuality, religion, and dis/ability, among all aspects of your identity, is your choice whether to disclose (e.g., should it come up in classroom conversation about our experiences and perspectives) and should be self-identified, not presumed or imposed. I will do my best to address and refer to all students accordingly, and I ask you to do the same for all of your fellow Terps.

## Communication with Instructor:

Email: If you need to reach out and communicate with me, please email me at nhassan@umd.edu. Please DO NOT email me with questions that are easily found in the syllabus or on ELMS (i.e. When is this assignment due? How much is it worth? etc.) but please DO reach out about personal, academic, and intellectual concerns/questions.

ELMS: I will send IMPORTANT announcements via ELMS messaging. You must make sure that your email & announcement notifications (including changes in assignments and/or due dates) are enabled in ELMS so you do not miss any messages. You are responsible for checking your email and Canvas/ELMS inbox with regular frequency.

## Communication with Peers:

With a diversity of perspectives and experiences, we may find ourselves in disagreement and/or debate with one another. As such, it is important that we agree to conduct ourselves in a professional manner and that we work together to foster and preserve a virtual classroom environment in which we can respectfully discuss and deliberate controversial questions.

I encourage you to confidently exercise your right to free speech—bearing in mind, of course, that you will be expected to craft and defend arguments that support your position. Keep in mind, that free speech has its limit and this course is NOT the space for hate speech, harassment, and derogatory language. I will make every reasonable attempt to create an atmosphere in which each student feels comfortable voicing their argument without fear of being personally attacked, mocked, demeaned, or devalued.

Any behavior (including harassment, sexual harassment, and racially and/or culturally derogatory language) that threatens this atmosphere will not be tolerated. Please alert me immediately if you feel

threatened, dismissed, or silenced at any point during our semester together and/or if your engagement in discussion has been in some way hindered by the learning environment.

## Grading Structure

| | |
|---|---|
| **Assignments** | **50%** |
| **Project** | **40%** |
| **Paper Presentation** | **10%** |
| **Total** | **100%** |

## Major Assignments

### Assignments

There will be a total of 5~6 assignments. These assignments are meant to assess your mastery of the topics and techniques covered in class. You may work with your colleagues to figure out the underlying concepts and problem-solving processes, but are expected to work individually to answer the specific problems that are assigned. Completed assignments will be submitted via Canvas/ELMS.

### Paper Presentation

You will need to present a research article during the course. It will carry 10% of the total grade. Details about the presentation format will be provided in the class.

### Project

You will work on a final project where you create an experiment, collect data, analyze it, and present your results. There will be a few milestones specific to the project, including a project proposal, a progress report (update), a presentation, and a final paper. Additional details here. https://docs.google.com/document/d/10WmHAXuF9OliuHNGdBlPHsJ6Jhh9xUBb/edit?usp=sharing&ouid=106552095934311980328&rtpof=true&sd=true.

Course Project (40%):

Proposal Presentation: 10%

Mid Update: 10%

Final Presentation: 10%

Deliverables: 10%

## Grades

All assessment scores will be posted on the course ELMS page. If you would like to review any of your grades (including the exams), or have questions about how something was scored, please email me to schedule a time for us to meet and discuss.

Late work will not be accepted for course credit so please plan to have it submitted well before the scheduled deadline. I am happy to discuss any of your grades with you, and if I have made a mistake I will immediately correct it. Any formal grade disputes must be submitted in writing and within one week of receiving the grade.

Final letter grades are assigned based on the percentage of total assessment points earned. To be fair to everyone I have to establish clear standards and apply them consistently, so please understand that being

close to a cutoff is not the same as making the cut (89.99 ≠ 90.00). It would be unethical to make exceptions for some and not others.

| Final Grade Cutoffs | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| + | 97.00% | + | 87.00% | + | 77.00% | + | 67.00% | + | |
| A | 94.00% | B | 84.00% | C | 74.00% | D | 64.00% | F | <60.0% |
| - | 90.00% | - | 80.00% | - | 70.00% | - | 60.00% | - | |

## Tentative Course Outline

| Date | Week# | Topic | Assignments/Projects |
|---|---|---|---|
| August 28, 2023 | 1 | Introduction to Computational Text Analysis. Elements of Text. Introduction to Google Colab. Text Tokenization using Python | A1: Release |
| September 11, 2023 | 2 | Regular Expression | A1: Due<br>A2: Release |
| September 18, 2023 | 3 | Guest Lecture by Professor Dr. Sarah Oates | |
| September 25, 2023 | 4 | Text Representation. Bag of Words Models. Word Embeddings | A2: Due |
| October 2, 2023 | 5 | Text Similarity. Jaccard Similarity. Euclidean Distance. Cosine Similarity | A3: Release |
| October 9, 2023 | 6 | Project Proposal Presentation in Class | Project Proposal Due |
| October 16, 2023 | 7 | Text Clustering. Kmeans Clustering. Cluster Evaluation and Visualization | A3: Due<br>A4: Release |
| October 23, 2023 | 8 | Text Classification. Nearest Neighbor. Support Vector Machine. | |
| October 30, 2023 | 9 | Topic Modeling and its Application | A4: Due |
| November 6, 2023 | 10 | Sentiment Analysis.<br>Project Update Presentation in Class | Project Update Due |
| November 13, 2023 | 11 | Parts of Speech Tagging. Named Entity Recognition. Pattern Matching. | A5: Release |
| November 20, 2023 | 12 | WordNet. Synonym, Homonym, Hyponym. FrameNet<br>Ethics in Text Analysis | A5: Due |
| November 27, 2023 | 13 | Information Extraction. Dependency Parsing. Coreference Resolution | |
| December 4, 2023 | 14 | Paper Presentation in Class | Paper Presentation Due |
| December 11, 2023 | 15 | Final Project Presentation in Class | Final Presentation Due |

Note: This is a tentative schedule, and subject to change as necessary – monitor the course ELMS page for current deadlines. In the unlikely event of a prolonged university closing, or an extended absence from the university, adjustments to the course schedule, deadlines, and assignments will be made based on the duration of the closing and the specific dates missed.

## Resources & Accommodations

### Accessibility and Disability Services

The University of Maryland is committed to creating and maintaining a welcoming and inclusive educational, working, and living environment for people of all abilities. The University of Maryland is also committed to the principle that no qualified individual with a disability shall, on the basis of disability, be excluded from participation in or be denied the benefits of the services, programs, or activities of the University, or be subjected to discrimination. The **Accessibility & Disability Service (ADS)** provides reasonable accommodations to qualified individuals to provide equal access to services, programs and activities. ADS cannot assist retroactively, so it is generally best to request accommodations several weeks before the semester begins or as soon as a disability becomes known. Any student who needs accommodations should contact me as soon as possible so that I have sufficient time to make arrangements.

For assistance in obtaining an accommodation, contact Accessibility and Disability Service at 301-314-7682, or email them at **adsfrontdesk@umd.edu**. Information about **sharing your accommodations with instructors**, **note taking assistance** and more is available from the **Counseling Center**.

### Student Resources and Services

Taking personal responsibility for your own learning means acknowledging when your performance does not match your goals and doing something about it. I hope you will come talk to me so that I can help you find the right approach to success in this course, and I encourage you to visit **UMD's Student Academic Support Services website** to learn more about the wide range of campus resources available to you.

In particular, everyone can use some help sharpening their communication skills (and improving their grade) by visiting **UMD's Writing Center** and schedule an appointment with the campus Writing Center.

You should also know there are a wide range of resources to support you with whatever you might need (**UMD's Student Resources and Services website** may help). If you feel it would be helpful to have someone to talk to, visit **UMD's Counseling Center** or **one of the many other mental health resources on campus**.

### Basic Needs Security

If you have difficulty affording groceries or accessing sufficient food to eat every day, or lack a safe and stable place to live, please visit **UMD's Division of Student Affairs website** for information about resources the campus offers you and let me know if I can help in any way.

### Technology Policy

Please refrain from using cellphones, laptops, and other electronic devices during class sessions unless we have designated such use as part of a class exercise.

### Course Evaluation

Please submit a course evaluation through CourseEvalUM in order to help faculty and administrators improve teaching and learning at Maryland. All information submitted to CourseEvalUM is confidential. Campus will notify you when CourseEvalUM is open for you to complete your evaluations for fall semester courses. Please go directly to the **Course Eval UM website** to complete your evaluations. By completing

all of your evaluations each semester, you will have the privilege of accessing through Testudo, the evaluation reports for the thousands of courses for which 70% or more students submitted their evaluations.

### Copyright Notice

Course materials are copyrighted and may not be reproduced for anything other than personal use without written permission.

## AI Use in Assignments

In this course, you are welcome to use artificial intelligence (AI)-powered programs such as ChatGPT or Bing AI to help you create outlines or first drafts of your work. These tools can be a great starting place, and I encourage you to spend time editing your final draft into something you are proud of. It is your responsibility to verify that any information you get in these drafts is accurate and represents your own point of view well. If you have any questions about this policy and what constitutes acceptable or unacceptable use of AI-based tools, please do not hesitate to ask me -- I would be happy to talk with you about this!