

Course Outline

Course Code	RSM 317 H1 F
Course Name	Text Mining and Natural Language Processing
Term, Year	Fall 2026
Course Meets	Wed, 11:00–13:00 Rotman Finance Lab
Web page URL	https://q.utoronto.ca

Instructor Details

Name	Email	Office Hours
Gerhard Trippen	Gerhard.Trippen@rotman.utoronto.ca	Wed, 15:00–17:00 (RT406); and by appointment
Yufeng Zheng (TA)	yufeng.zheng@rotman.utoronto.ca	Contact by email for appointments

Course Description

This course introduces a broad range of text-mining and natural-language-processing (NLP) techniques built on machine learning. These methods identify and quantify structure in text data to answer business questions and generate managerial insight, with an emphasis on model validation and clear communication of results. The course follows a "white-box" approach: rather than treating models as black boxes, you will develop an understanding of the underlying algorithmic and statistical structures and how they apply to text. It is a hands-on, Python-based course that applies these techniques to real-world, large-scale text data. Because today's generative-AI systems are themselves built on the methods studied here, the course also treats the responsible, well-understood use of such tools—verifying and taking ownership of their output—as a core professional skill.

Learning Outcomes

By the end of this course, students will be able to:

- Apply text-preprocessing techniques—tokenization, normalization, regular expressions, and spell-checking—to prepare raw, unstructured text for analysis.
- Engineer text features using Bag-of-Words, TF-IDF, and document-term matrix representations, and explain how each representation shapes downstream models.
- Build and evaluate text-classification models, including neural networks in Keras, for tasks such as sentiment analysis and categorization.
- Apply clustering and topic-modelling techniques (e.g., k-means, LSI, and LDA) to discover latent structure, such as document groupings and topics, in large text collections.
- Measure document similarity and word meaning using vector-space and word-embedding methods such as Word2Vec.
- Interpret model output to derive and communicate actionable managerial insights from unstructured text for non-technical stakeholders.

- Validate text-based machine-learning models and communicate their results and limitations effectively.
- Use Python libraries (scikit-learn, NLTK, and Keras) to implement text-mining and NLP workflows on real-world data.
- Critically evaluate, debug, and verify AI-generated code and output, taking full responsibility for its correctness—a core AI-literacy skill for applying generative AI responsibly.

Course Prerequisites

CSC108H1 Introduction to Computer Programming.

Course Materials

Required Readings

<https://www.datacamp.com/>

DataCamp (online, interactive Python courses). A free six-month DataCamp subscription is provided to every enrolled student; see Course Format and Expectations below.

Recommended / Optional Readings

The following textbooks are optional references, available through the UofT Library or online:

- Dipanjan Sarkar (2019), Text Analytics with Python: A Practitioner's Guide to Natural Language Processing, 2nd ed. — primary reference for the weekly readings.
- B. Bengfort, R. Bilbro & T. Ojeda, Applied Text Analysis with Python (O'Reilly).
- Wes McKinney (2022), Python for Data Analysis, 3rd ed. — free online at <https://wesmckinney.com/book/>

Electronic Course Materials

This course will use the following electronic course materials:

- DataCamp: <https://www.datacamp.com/>
- UofT JupyterHub: <https://jupyter.utoronto.ca/>
- Anaconda: <https://www.anaconda.com/>

These materials will cost a total of **\$0.00**. The use of these materials complies with all University of Toronto policies which govern fees for course materials.

All students will receive a free DataCamp subscription for six months, with access to all the online courses DataCamp offers (not just the chapters covered in the course). To register for DataCamp, an email address is required, which you will share with the instructor at the beginning of the term. Students who prefer to opt out and not use DataCamp may submit solutions to exercises to the instructor by email.

Evaluation and Grades

Grades are a measure of the knowledge and skills developed by a student within individual courses. Each student will receive a grade on the basis of how well they have command of the course materials, skills and learning objectives of the course.

Work	Percentage of grade	Due Date
In-Class Exercises (best 10 of 12)	10%	Ongoing (in class)
Group Assignment 1 SEC 10-K / 10-Q Parsing	12%	Fri, 2026-10-23 23:59
Group Assignment 2 Sentiment Analysis (Keras ANN + prediction)	16%	Fri, 2026-11-20 23:59

Group Assignment 3 Clustering & Topic Modelling	12%	Fri, 2026-12-04 23:59
Final Examination (in person)	50%	TBD (Dec 10–22 exam period)
Total	100%	

Course Format and Expectations

Classes are held in the Rotman Finance Lab and combine concept explanation with hands-on work in Jupyter. You are expected to work through coding examples during class. The assignments and in-class exercises are designed to build the fluency you will need for the in-person final exam.

In-Class Exercises

In-class exercises are short, individual questions posed during the session that test your understanding of the material as it is presented. Some are conceptual; others refer to something generated live in class, such as a code output, a chart, or a diagram. Because they depend on what happens in the room, they are recorded in a Quercus quiz that is open only during the class session and cannot be completed remotely. These exercises are completed without generative-AI assistance and serve as ongoing, low-stakes practice for the in-person final examination. You receive partial marks for a genuine attempt and full marks for correct answers. Only your best 10 of 12 sessions count toward the final grade, so an occasional missed class is automatically accommodated; for this reason, no special consideration is granted for missed in-class exercises.

Writing Assignments or Presentations

Your group assignments are intended to help you develop your communication skills. How well you communicate your ideas, in writing or orally, will be considered in the evaluation of the assignment. In your written assignments, you should aim for clarity, strong organization, concision, professionalism, and correct grammar. Your presentations should reflect strong planning and organization, clarity of speech, and an engaging demeanour. Sources, whether in written or presentation assignments, should always be correctly attributed.

Support is available through the RC Centre for Professional Skills (CPS) for students who would like help or feedback on their writing or speaking (presentations). CPS offers both individual and group appointments with trained writing instructors and presentation coaches who are familiar with the RC program and common types of business assignments. Coaches can also help you effectively work with generative AI in your work, where appropriate. You can also access your college Writing Centres for help with written assignments.

You can book an appointment [with a writing or presentation coach](#) through the RC Centre for Professional Skills Writing Centre. For more information about writing centres, student supports, and study resources, see the [Writing and Presentation Coaching academic support page](#).

Team or Group Assignments

The group assignments require students to work in teams of 3–4 (preferably 4). Learning to work together in teams is a crucial transferrable skill you will use not only in your coursework, but also in your future careers. Support is available if you encounter common teamwork challenges such as:

- Team members feeling left out of the team.
- Team members not responding in a timely manner to communication.
- Division or quality of work among team members being unequal or unfair.

Consult the [Centre for Professional Skills Teamwork Resources page](#) for tips, strategies, and best practices. You can also [book an appointment with a teamwork mentor](#) through the RC

Centre for Professional Skills Writing Centre or by emailing rotmancommerce.teamworkhelp@utoronto.ca. Teamwork mentors can help you resolve or mitigate conflict, strategize on planning, or improve team communication.

If you are a student registered with Accessibility Services, and extensions are one of your academic accommodations, consult with your Accessibility Advisor about the teamwork in this course.

Rotman Commerce Attendance Policy

Rotman Commerce students are expected to make every effort to attend each class. Infrequently, students may miss term work, e.g., quizzes, assignments, etc., due to unplanned and extenuating circumstances and must follow the Request for Special Consideration process as outlined in their course outlines. However, Rotman Commerce will not approve any Request for Special Consideration for participation marks for missed classes. Any such request will be denied.

Missed Tests and Assignments

Students who miss a term test or assignment for reasons entirely beyond their control (e.g. illness) may request special consideration **within 2 business days** of the missed midterm/test/assignment due date.

In such cases, students must:

1. Complete the Request for Special Consideration form: <https://uoft.me/RSMConsideration>
2. Provide documentation to support the request, eg. Absence Declaration from [ACORN](#), [Verification of Illness Form](#) etc.

Please note: Students may use the Absence Declaration on ACORN ***one time per term*** to report an absence and request consideration. **Any subsequent absence will require a Verification of Illness form or other similar relevant documentation.**

Students who do not submit their requests and documentation within 2 days may receive a grade of 0 (zero) on the missed course deliverable.

If a Request for Special Consideration is approved, the weight of a missed group assignment will normally be transferred to the final examination; as a result no single component will exceed 80% of the final grade. Because the in-class coding exercises are graded on a best-10-of-12 basis, missed individual exercises are already accounted for and are not eligible for re-weighting or special consideration.

Final Exams: If you miss the final exam in this course for a legitimate reason (illness, etc) you will need to contact your College Registrar to file a petition for a deferred exam. This deferred exam will be written at a later date as established by the Faculty of Arts & Science. Instructions can be found here: <https://www.artsci.utoronto.ca/current/faculty-registrar/petitions-appeals/preparing-petition>

Late Assignments

All assignments are due on the date and at the time specified in Quercus. Late submissions will normally be penalized by 20% if the assignment is not received on the specified date, at the specified time. A further penalty of 10% will be applied for each subsequent day.

Statement on Equity, Diversity and Inclusion

The University of Toronto is committed to equity, human rights and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect where all members of our community can express themselves, engage with each other, and respect one another's differences. U of T does not condone discrimination or harassment against any persons or communities.

Commitment to Accessibility

The University is committed to inclusivity and accessibility, and strives to provide support for, and facilitate the accommodation of, individuals with disabilities so that all may share the same level of access to opportunities and activities offered at the University.

If you require accommodations for a temporary or ongoing disability or health concern, or have any accessibility concerns about the course, the classroom or course materials, please [email Accessibility Services](#) or visit the [Accessibility Services website](#) for more information as soon as possible. Obtaining your accommodation letter may take up to several weeks, so get in touch with them as soon as possible. If you have general questions or concerns about the accessibility of this course, you are encouraged to reach out to your instructor, course coordinator, or Accessibility Services.

Plagiarism Detection

Normally, students will be required to submit their course essays to the University's plagiarism detection tool for a review of textual similarity and detection of possible plagiarism. In doing so, students will allow their essays to be included as source documents in the tool's reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of this tool are described on the [University's Plagiarism Detection Tool FAQ](#) page from Centre for Teaching Support & Innovation.

In addition, source code submitted for the group assignments may be reviewed using **MOSS** (Measure of Software Similarity), an automated system for detecting similarity among programs, to help identify possible plagiarism in submitted code.

Generative AI / ChatGPT

Generative-AI tools (ChatGPT, Claude, Copilot, and similar assistants) are permitted on the work you complete outside of class — the group assignments (and any individual coding assignments) — subject to the conditions below. They are not permitted on the in-class exercises or the final examination. Used well, these tools let you push your work further and spend your time on interpretation and judgement rather than boilerplate, and learning to use them responsibly is itself part of what this course develops.

Disclosure. When you use a generative-AI tool on any submitted assignment, you must disclose it in a brief appendix or note at the end of the deliverable, stating which tool(s) you used and for which parts of the work — for example, "Used ChatGPT to draft an initial version of the Q3 code; verified, tested, and edited by hand." Undisclosed use of generative AI on graded work may be treated as an unauthorized aid under the Code of Behaviour on Academic Matters.

Responsibility. You are accountable for everything you submit, including any AI-generated content. These tools readily produce plausible-looking output that is wrong, biased, or fabricated, so you are expected to verify, test, and understand any code or analysis you turn in. Verifying and taking ownership of AI output is itself a graded skill in this course; "the AI gave me that answer" is not an acceptable defence for errors in your submission.

In-class exercises and final examination. Generative-AI tools may not be used on the in-class exercises or the final examination. However, both are open-book: you may use the lecture materials and any code written in class or in your own assignments, but not generative-AI tools, the internet, or any other outside assistance. The final is written in person, and your answers on all of these must reflect your own individual understanding of the material.

A note in your own interest. The assignments you may use AI on are your practice for a final examination you will write with no AI tools at all. You may use AI on that practice — but be warned: if you rely on it to produce work you do not fully understand, you will save a little time now and arrive at the final unable to do the work under your own power. Students who lean on AI throughout the term regularly underperform on in-person assessments for exactly this reason. The assignments are also worth less, in total, than the final they prepare you for, so trading real practice for an easy grade now is a poor trade. Use these tools to accelerate work you understand — not to bypass the learning the course is designed to build.

Private study and access. Separately, you may always use generative AI as a private learning aid — to explain a concept or test your understanding, in the same way you would consult a textbook, tutor, or peer — and such personal study use does not need to be disclosed. Some generative-AI tools and course platforms require accounts or paid subscriptions; if you have concerns about cost, privacy, or security, contact the instructor about an opt-out or alternative.

Academic Integrity

Academic Integrity is a fundamental value essential to the pursuit of learning and scholarship at the University of Toronto. Participating honestly, respectfully, responsibly, and fairly in this academic community ensures that the U of T degree that you earn will continue to be valued and respected as a true signifier of a student's individual work and academic achievement. As a result, the University treats cases of academic misconduct very seriously.

[The University of Toronto's Code of Behaviour on Academic Matters](#) outlines the behaviours that constitute academic misconduct, the process for addressing academic offences and the penalties that may be imposed. You are expected to be familiar with the contents of this document. Potential offences include, but are not limited to:

In papers and assignments

- Using someone else's ideas or words without appropriate acknowledgement.
- Submitting your own work in more than one course without the permission of the instructor.
- Making up sources or facts.
- Obtaining or providing unauthorized assistance on any assignment (this includes collaborating with others on assignments that are supposed to be completed individually).

On test and exams

- Using or possessing any unauthorized aid, including a cell phone.
- Looking at someone else's answers.
- Misrepresenting your identity.
- Submitting an altered test for re-grading.

Misrepresentation

- Falsifying institutional documents or grades.
- Falsifying or altering any documentation required by the University, including (but not limited to) medical notes.

All suspected cases of academic dishonesty will be investigated by the procedures outlined in the [Code of Behaviour on Academic Matters](#). If you have any questions about what is or is not permitted in the course, please do not hesitate to contact the course instructor. If you have any questions about appropriate research and citation methods, you are expected to seek out additional information from the instructor or other U of T or RC resources such as the RC Centre for Professional Skills, the College Writing Centres or the Academic Success Centre.

Email

At times, the course instructor may decide to communicate important course information by email. As such, all U of T students are required to have a valid UTmail+ email address. You are responsible for ensuring that your UTmail+ email address is set up and properly entered on ACORN. For more information visit the [Information Commons Help Desk](#).

Forwarding your utoronto.ca email to a Gmail or other type of email account is not advisable. In some cases, messages from utoronto.ca addresses sent to Gmail accounts are filtered as junk mail, which means that important messages from your course instructor may end up in your spam or junk mail folder.

Recording Lectures

Lectures and course materials prepared by the instructor are considered by the University to be an instructor's intellectual property covered by the Canadian Copyright Act. Students wishing to record a lecture or other course material in any way are required to ask the instructor's explicit permission and may not do so unless permission is granted. Students who have been previously granted permission to record lectures as an accommodation for a disability are excepted. This includes tape recording, filming, photographing PowerPoint slides, Quercus materials, etc.

If permission for recording is granted by the instructor (or via Accessibility Services), it is intended for the individual student's own study purposes and does not include permission to "publish" them in any way. It is forbidden for a student to publish an instructor's notes to a website or sell them in any other form without formal permission.

Weekly Schedule

Sessi on	Date	Topic	Readings (Sarkar, 2019) / Notes
1	2026-09-09	Introduction & Overview; Working with Text Data	Ch. 1: NLP Basics
2	2026-09-16	Python Foundations for Text Analysis	Ch. 2: Python for NLP
3	2026-09-23	Processing & Understanding Text	Ch. 2–3
4	2026-09-30	Regular Expressions & Text Preprocessing	Ch. 2–3
5	2026-10-07	Tokenization, Normalization & Spell-Checking	Ch. 3
6	2026-10-14	Feature Engineering: Bag-of-Words & TF-IDF	Ch. 4
7	2026-10-21	Text Classification with Neural Networks	Ch. 5; Ch. 9 GA1 due Fri 2026-10-23
—	2026-10-28	Fall Reading Week — no class	—
8	2026-11-04	Sentiment Analysis with Deep Learning	Ch. 5; Ch. 9
9	2026-11-11	Text Similarity & Clustering	Ch. 7
10	2026-11-18	Topic Modelling (LSI, LDA)	Ch. 6 GA2 due Fri 2026-11-20
11	2026-11-25	Word Embeddings: Word2Vec	Ch. 4 (Advanced)
12	2026-12-02	Text Generation (GenAI) & Transformers	Ch. 10 GA3 due Fri 2026-12-04

Please note that the last day you can drop this course without academic penalty is Tuesday, November 17, 2026



Other Useful Links

- [Become a volunteer note taker](#)
- [Accessibility Services Note Taking Support](#)
- [Credit / No-Credit in RSM courses](#)
- [Rotman Commerce Academic Support](#)
- [Where to find teaching assistant opportunities](#)

URL links for print

- ACORN: <http://www.acorn.utoronto.ca/>
- Email Accessibility Services: accessibility.services@utoronto.ca
- Accessibility Services website: <http://studentlife.utoronto.ca/as>
- University's Plagiarism Detection Tool FAQ: <https://uoft.me/pdt-faq>
- The University of Toronto's Code of Behaviour on Academic Matters: <http://www.governingcouncil.utoronto.ca/policies/behaveac.htm>
- Information Commons Help Desk: <http://help.ic.utoronto.ca/category/3/utmail.html>
- Become a volunteer note taker: <https://studentlife.utoronto.ca/program/volunteer-note-taking/>
- Accessibility Services Note Taking Support: <https://studentlife.utoronto.ca/service/note-taking-support/>
- Credit / No-Credit in RSM courses: <https://rotmancommerce.utoronto.ca/current-students/degree-requirements/credit-no-credit-option/>
- Rotman Commerce Academic Support: <https://rotmancommerce.utoronto.ca/current-students/academic-support/>
- Book an appointment with a writing or presentation coach: <http://uoft.me/writingcentres>
- Writing and Presentation Coaching academic support page: <https://rotmancommerce.utoronto.ca/current-students/academic-support/writing-and-presentation-coaching/>
- Centre for Professional Skills Teamwork Resources page: <https://rotmancommerce.utoronto.ca/teamwork-resources>
- Book an appointment with a Teamwork Mentor: <http://uoft.me/writingcentres>