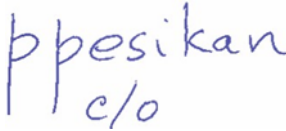


Course Outline

School:	Eng. Tech. & Applied Science
Department:	Information and Communication Engineering Technology (ICET)
Course Title:	Natural Language Processing and Recommender Systems
Course Code:	COMP 262
Course Hours/Credits:	56
Prerequisites:	COMP 247, COMP 258, COMP 254
Co-requisites:	N/A
Eligible for Prior Learning, Assessment and Recognition:	Yes
Originated by:	Mayy Habayeb
Creation Date:	Winter 2022
Current Semester:	Winter 2023
Approved by:	
	<hr/> Chairperson/Dean

Students are expected to review and understand all areas of the course outline.

Retain this course outline for future transfer credit applications. A fee may be charged for additional copies.

This course outline is available in alternative formats upon request.

Acknowledgement of Traditional Lands

Centennial is proud to be a part of a rich history of education in this province and in this city. We acknowledge that we are on the treaty lands and territory of the Mississaugas of the Credit First Nation and pay tribute to their legacy and the legacy of all First Peoples of Canada, as we strengthen ties with the communities we serve and build the future through learning and through our graduates. Today the traditional meeting place of Toronto is still home to many Indigenous People from across Turtle Island and we are grateful to have the opportunity to work in the communities that have grown in the treaty lands of the Mississaugas. We acknowledge that we are all treaty people and accept our responsibility to honor all our relations.

Course Description

The first part of this course focuses on natural language processing concepts and techniques. Topics include sentiment analysis, summarization, dialogue state tracking, etc. Students will apply these concepts to build a conversational interface (chat bot).

The second part of the course introduces recommender systems for predicting user preferences. Topics include the most fundamental techniques used in recommender systems, such as association rules and collaborative filtering. More advanced techniques will be also explained.

Program Outcomes

Successful completion of this and other courses in the program culminates in the achievement of the Vocational Learning Outcomes (program outcomes) set by the Ministry of Colleges and Universities in the Program Standard. The VLOs express the learning a student must reliably demonstrate before graduation. To ensure a meaningful learning experience and to better understand how this course and program prepare graduates for success, students are encouraged to review the Program Standard by visiting <http://www.tcu.gov.on.ca/pepg/audiences/colleges/progstan/>. For apprenticeship-based programs, visit <http://www.collegeoftrades.ca/training-standards>.

Course Learning Outcomes

The student will reliably demonstrate the ability to:

1. Evaluate the wide spectrum of problem statements, tasks, and solution approaches within Natural Language Processing (NLP).
2. Compare the various available techniques for text acquisition, cleaning, representation and feature engineering.
3. Explore various algorithms (Machine Learning/Deep Learning) and approaches for NLP product tasks, datasets, and pipelines.
4. Design, code & test different NLP applications using machine learning and deep learning methods to solve text classification problems related to sentiment analysis and news classification.
5. Design, code & test different NLP applications using machine learning and deep learning methods to solve Information extraction problems.
6. Design, code & test different NLP applications using machine learning and deep learning methods to solve conversational agent problems, such as chatbots using api's.
7. Fine-tune NLP solution based on the business problem and industry vertical.
8. Produce software solutions following best practices around release, deployment, and DevOps for

NLP systems.

9. Discuss and understand the taxonomy of a recommender system.
10. Examine different types of recommender algorithms association rules, content based, collaborative filtering, hybrid.
11. Explore the use cases for Knowledge-based recommender systems.
12. Evaluate the use cases of Non-personalized recommender systems.
13. Design/code and test a simple recommender system to solve a real business problem.

Essential Employability Skills (EES)

The student will reliably demonstrate the ability to*:

1. Communicate clearly, concisely and correctly in the written, spoken, and visual form that fulfills the purpose and meets the needs of the audience.
2. Respond to written, spoken, or visual messages in a manner that ensures effective communication.
3. Execute mathematical operations accurately.
4. Apply a systematic approach to solve problems.
5. Use a variety of thinking skills to anticipate and solve problems.
7. Analyze, evaluate, and apply relevant information from a variety of sources.
9. Interact with others in groups or teams in ways that contribute to effective working relationships and the achievement of goals.
10. Manage the use of time and other resources to complete projects.
11. Take responsibility for one's own actions, decisions, and consequences.

**There are 11 Essential Employability Skills outcomes as per the Ministry Program Standard. Of these 11 outcomes, the following will be assessed in this course.*

Global Citizenship and Equity (GC&E) Outcomes

N/A

Methods of Instruction

Engaging and interactive lecture content.

Lab demonstrations and tutorials.

Hands on practical lab exercises.

Interactive discussion forms and boards.

Team project.

Text and other Instructional/Learning Materials

Text Book(s):

1-Practical Natural Language Processing by Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, and Harshit Surana published by O'Reilly Media, Inc. 2020.

2-Practical Natural Language Processing with Python : With Case Studies from Industries Using Text Data at Scale by Mathangi Sri published by Apress 2020.

3-Practical Recommender Systems by Kim Flak published by Manning Publications 2019

4-Hands-On Recommendation Systems with Python by Rounak Banik published by Packt publishing 2018

Online Resource(s):

https://learning.oreilly.com/library/view/practical-natural-language/9781492054047/?sso_link=yes&sso_link_from=centennial-college
<https://learning.oreilly.com/library/view/practical-natural-language/9781484262467/?ar=>
<https://learning.oreilly.com/library/view/practical-recommender-systems/9781617292705/>
<https://learning.oreilly.com/library/view/hands-on-recommendation-systems/9781788993753/a894116c-3c71-4e33-b46f-2e04ff05162c.xhtml>

Evaluation Scheme

- ✦ Lab Assignments: Three lab assignments
 - 1-NLP-Text classification to solve a real business problem
 - 2-NLP-Conversational agent using APIs for a real business problem
 - 3-Recommender system to solve a real business problem
- ✦ Quizzes: Two quizzes
- ✦ Test #1: Midterm Test (Theory and hands-on)
- ✦ Project: Group project "Product recommender system using NLP"
- ✦ Online interaction and discussion boards: Contributing to On-line discussion boards and participating in virtual classroom and labs

Evaluation Name	CLO(s)	EES Outcome(s)	GCE Outcome(s)	Weight/100
Lab Assignments	3, 4, 5, 6, 13	1, 2, 5, 7, 9, 11		24
Quizzes	1, 2, 3, 9, 10, 11, 12	2, 4, 5, 10, 11		16
Test #1	1, 2, 3, 4, 5, 7, 8	1, 3, 4, 5, 7, 10, 11		25
Project	1, 2, 3, 4, 5, 10, 11, 12, 13	1, 2, 4, 5, 7, 9, 10, 11		25
Online interaction and discussion boards	1, 2, 7, 8, 9, 10, 11, 12	1, 2, 7, 11		10
Total				100%

If students are unable to write a test they should immediately contact their professor or program Chair for advice. In exceptional and well documented circumstances (e.g. unforeseen family problems, serious illness, or death of a close family member), students may be able to write a make-up test.

All submitted work may be reviewed for authenticity and originality utilizing Turnitin®. Students who do not wish to have their work submitted to Turnitin® must, by the end of the second week of class, communicate this in writing to the instructor and make mutually agreeable alternate arrangements.

When writing tests, students must be able to produce official Centennial College photo identification or they may be refused the right to take the test or test results will be void.

Tests or assignments conducted remotely may require the use of online proctoring technology where the student's identification is verified and their activity is monitored and/or recorded, both audibly and visually through remote access to the student's computer and web camera. Students must communicate in writing to the instructor as soon as possible and prior to the test or assignment due date if they require an alternate assessment format to explore mutually agreeable alternatives.

Student Accommodation

The Centre for Accessible Learning and Counselling Services (CALCS) (<http://centennialcollege.ca/calcs>) provides programs and services which empower students in meeting their wellness goals, accommodation and disability-related needs. Our team of professional psychotherapists, social workers, educators, and staff offer brief, solution-focused psychotherapy, accommodation planning, health and wellness education, group counselling, psycho-educational workshops, adaptive technology, and peer support. Walk in for your first intake session at one of our service locations (Ashtonbee Room L1-04, Morningside Room 190, Progress Room C1-03, The Story Arts Centre Room 285, Downsview Room 105) or contact us at calcs@centennialcollege.ca, 416-289-5000 ext. 3850 to learn more about accessing CALCS services.

Use of Dictionaries

- Any dictionary (hard copy or electronic) may be used in regular class work.

Program or School Policies

N/A

Course Policies

Late lab assignments will be accepted for 3 days after the due date has passed, subject to the following penalty structure:

A 10% penalty per day of lateness.

Assignments will not be accepted past the 3 day limit and will not earn any marks. There are no extensions for discussion boards.

College Policies

Students should familiarize themselves with all College Policies that cover academic matters and student conduct.

All students and employees have the right to study and work in an environment that is free from discrimination and harassment and promotes respect and equity. Centennial policies ensure all incidents of harassment, discrimination, bullying and violence will be addressed and responded to accordingly.

Academic Honesty

Academic honesty is integral to the learning process and a necessary ingredient of academic integrity. Forms of academic dishonesty include cheating, plagiarism, and impersonation, among others. Breaches of academic honesty may result in a failing grade on the assignment or course, suspension, or expulsion from the college. Students are bound to the College's AC100-11 Academic Honesty and Plagiarism policy.

To learn more, please visit the Libraries information page about Academic Integrity

<https://libraryguides.centennialcollege.ca/academicintegrity> and review Centennial College's Academic Honesty Module:

https://myappform.centennialcollege.ca/centennial/articulate/Centennial_College_Academic_Integrity_Module_%202/story.html

Use of Lecture/Course Materials

Materials used in Centennial College courses are subject to Intellectual Property and Copyright protection, and as such cannot be used and posted for public dissemination without prior permission from the original creator or copyright holder (e.g., student/professor/the College/or third-party source). This includes class/lecture recordings, course materials, and third-party copyright-protected materials (such as images, book chapters and articles). Copyright protections are automatic once an original work is created, and applies whether or not a copyright statement appears on the material. Students and employees are bound by College policies, including AC100-22 Intellectual Property, and SL100-02 Student Code of Conduct, and any student or employee found to be using or posting course materials or recordings for public dissemination without permission and/or inappropriately is in breach of these policies and may be sanctioned.

For more information on these and other policies, please visit www.centennialcollege.ca/about-centennial/college-overview/college-policies.

Students enrolled in a joint or collaborative program are subject to the partner institution's academic policies.

PLAR Process

This course is eligible for Prior Learning Assessment and Recognition (PLAR). PLAR is a process by which course credit may be granted for past learning acquired through work or other life experiences. The PLAR process involves completing an assessment (portfolio, test, assignment, etc.) that reliably demonstrates achievement of the course learning outcomes. Contact the academic school to obtain information on the PLAR process and the required assessment.

This course outline and its associated weekly topical(s) may not be reproduced, in whole or in part, without the prior permission of Centennial College.

Semester:	Winter 2023	Professor Name:	see e-centennial
Section Code:	All	Contact Information:	see e-centennial
Meeting Time & Location:	see my-centennial	Delivery Method:	Section 001 - 003 - On-site, 401- Online

Topical Outline (subject to change):

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
1	Course Overview. Introduction to NLP. Challenges of NLP. NLP pipelines.	Module #1 Material Chapter 1 & Chapter 2 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder)	Compare various NLP tasks and their challenges and use cases. Discuss the main blocks of language (context, syntax, morphemes and lexemes, phonemes) and how they relate to NLP tasks and applications. Examine the challenges in NLP. Explain the main AI approaches to NLP (Heuristic, ML, DL) and identify the most common ML/DL algorithms used for NLP. Discuss the main steps in the NLP pipeline. Evaluate the various "Data Augmentation" Techniques to create more text data.	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Discussion board #1 made available	
2	Advanced pre-processing NLP feature engineering. Metrics. Text Representation. Word embedding.	Module #2 material Chapter 2 & Chapter 3 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder)	Evaluate text preprocessing techniques. Discuss regular expressions and their use in pre-processing. Evaluate basic text representations. Examine the concept of word embedding and the most popular word embedding models. Compare possible approaches using word embedding. Design, code and test an NLP sentiment analysis model.	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Assignment #1 made available	
3	Text classification. Lexicon-based sentiment analysis. ML for text classification (Naïve Bayes, SVM, Logistic regression) DL for text classification. BERT model.	Module #3 Material Chapter 4 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder) Chapter 3(Sri)	Discuss the definition of text classification and the various applications of text classification. Examine the different approaches to text classification. Examine "Active Learning" approaches to text classification. Explore the concept of "Sentiment analysis". Design, code and test text classifiers and sentiment analysis models, using Lexicon, ML and DL approaches. (Lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Discussion board #1 due Discussion board #2 made available	

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
4	Information extraction	Module #4 material. Chapter 5 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder)	Discuss the definition of Information extraction (IE) and list the various use cases for IE. Discuss the various Information extraction tasks. Examine the task of Named entity recognition (NER) and the Conditional random fields (CRFs). Examine the IE task of keyword and phrase extraction and approaches for implementation. Design, code, and test a named entity recognizer. (Lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Quiz #1 Discussion board #2 due	week #4
5	Conversational agents Introduction to chatbots and their history.	Module #5 material Chapter 6 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder)	List the various use cases of Chatbots. Evaluate the chatbot APIs available on the market. Compare the two types of chatbots: goal oriented and chitchats. Examine the Pipeline for Building Dialog Systems of a chatbot. Design and build a chatbot system using API. (Lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Assignment#1 due Assignment #2 made available	week #5
6	Chatbots algorithms Chatbots public datasets	Module #6 material Chapter 6 (Anuj Gupta; Harshit Surana; Sowmya Vajjala; Bodhisattwa Majumder)	Discuss various algorithms/tools used to build chatbots. Explore various datasets publicly available for dialog systems. Discuss the Sequence-to-sequence (seq2seq) models. Examine the Rasa Framework for chatbots. Design and build parts of a chatbot system using ML algorithms and public datasets (Lab).	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards		
7	Midterm test Project phase #1 presentations	Weeks 1-6	N/A	Test theory and hands-on. Project class presentations.	Test #1 Project phase #1 presentation.	Week #7
8	Recommender systems Introduction Similarities	Module#8 material Chapters 1 & 7 Falk Chapters 2 Rounak	Discuss the definition of Recommender systems and list the various use cases for recommender systems. Examine the taxonomy of a recommender system.	Lecture Demonstration Lab Session Class Discussion Videos	Discussion board #3 made available	

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
			Differentiate between a recommendation and a predication. Compare ways to measure similarities. Design code and test a simple recommender system (lab).	Discussion boards		
9	Non-Personalized recommendations Charts Items Frequency sets Association rules	Module #9 material Chapters 5 Falk Chapters 3 Rounak	Explore non-personalized recommender systems. Examine association rules algorithms. Design code and test a recommender systems using association rules. (lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Assignment #2 due Discussion board #3 due	
10	Content based filters.	Module#10 material Chapters 10 & 11 Falk Chapter 3 Rounak	Discuss knowledge-based recommenders. Examine content based filter recommender systems and their applications. Explain the "Latent Dirichlet Allocation" algorithm and its use in recommender engines. Design, code & test a Knowledge based recommender system. (lab) Design, code & test a content-based recommender systems using text description data. (lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Quiz#2 Assignment #3 made available Discussion board #4 made available	Week #10
11	Collaborative filtering. User based filtering. Item based filtering. Cold Strat problem. SVD.	Module#11 material. Chapters 2,7 & 8 Falk	Examine the theory behind collaborative based filter recommender systems. Compare user based to item based approaches and the utility of each. Investigate ways to calculate the predicated rating. Explore implementation strategies of collaborative filtering online vs offline. Examine challenges of collaborative filtering. Design, code and test a simple collaborative filtering recommender system. (Lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards		
12	Matrix-factorization Hybrid filters	Module#12 material. Chapters 4 & 5,8,11,12 Falk	Explore the use of matrix-factorization techniques with collaborative filtering. Examine the theory behind hybrid based filter recommender systems. Design, code, and test a model based recommender system. (Lab)	Lecture Demonstration Lab Session Class Discussion Videos Discussion boards	Assignment #3 due Discussion board #4 due	
13	Testing recommender	Module#13 material.	Examine ways to evaluate and test recommender systems.	Lecture Demonstration		

Week	Topics	Readings/Materials	Weekly Learning Outcome(s)	Instructional Strategies	Evaluation Name and Weight	Evaluation Date
	systems. Review.			Lab Session Class Discussion Videos Discussion boards		
14	Project presentations	Module materials.	Implementation of a product recommender systems utilizing NLP	Class presentation and discussions	Project presentation	week 14