Course Description

This is an online and not a lecture-based course. Students will engage in a hands-on application of big data analytics methods and techniques in a project setting. The capstone course is intended to bring traditional data analytics methods, techniques, and approaches and apply them to a real-world area of interest. Working with a faculty supervisor, students will apply what they have learned in the prerequisite courses and implement their analysis through Python or R platforms.

Course Focus and Scope

As the final capstone course of the Data Analytics, Big Data, and Predictive Analytics certificate program, the course addresses different aspects of data analytics. It may tackle problems from various perspectives and use traditional machine learning methods such as classification, clustering and knowledge induction. The course aims to provide a practical foundation in the applied data analytics domain. An individually designed and implemented project will be delivered, showing skills learned to (re)solve a real-world problem.

Course Learning Outcomes

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

- Apply traditional machine learning techniques to frame a research problem in a way that could be addressed using am ML-based approaches
- Use relevant and visualization R packages or Python libraries to address the problem under study
- Replicate state-of-the-art studies and highlight the similarities and differences between the actual and replicated work in terms of applied datasets, approaches, tools, and outcomes
- Build a solid foundation on “how” and “when” to apply an ML-based approach
- Communicate the findings of the research effectively in both written and oral presentations

Teaching Methods

Each student will be granted access to a Google Cloud Platform equipped with RStudio and Python Jupyter Notebook to complete their capstone project. These platforms will be available until the end of the term and will be archived by then. Students are required to follow up with their designated supervisors to ensure the validity of their proposed workflow through one-on-one online meetings.

Course Schedule

Module 1

Project Abstract

Topics

- Identifying a theme and research questions
- Selecting a dataset
- Writing the project abstract

Learning Objectives

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

- Identify a theme to run their analyses from themes that have been taught in the pre-requisite courses (Text Classification and Sentiment Analysis, Predictive Analytics, Recommender System Development, Data Mining and Knowledge Discovery)
- Conduct research to select a publicly available dataset for their project (either from a list of repositories suggested by the course lead or on their own)
- Formulate at least three research questions recognizing the main problems of the selected theme (Notably, the research questions should be connected, relevant, and justifying the research effort.)

Assessments

- Project Abstract (5% of the final grade), due by the beginning of Week 3

Module 2

Literature Review, Data Description, and Project Approach

Topics

- Defining the problem under study
- Describing the working data set and any imposed constraints
- Relating the selected topic to the body of knowledge
- Comparing the proposed methodology with what other practitioners/researchers did in the past

Learning Objectives

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

- Explain the problem under study
- Comment on what has been already conducted and analyzed concerning the selected theme
- Specify how their proposed work fits in with past research according to the literature review
- List the implication of answering the tentative research questions

Assessments

- Literature Review, Data Description, and Approach (25% of the final grade), due by the beginning of Week 6

Module 3

Initial Results and the Code

Topics

- Data analyses, including summarizing and visualizing data
- Data preparation, including selecting, preprocessing, and transforming data
- Algorithms evaluation, including testing options, exploring algorithms, and reporting results

Learning Objectives

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

- Provide a proof of concept ensuring the feasibility of the proposed solution
- Summarize and visualize data features
- Format, clean, sample, scale, decompose, or aggregate data
- Build and evaluate at least one model answering question
- Document and check in their codes using the GitHub repository

Assessments

- Initial Results and the Code (10% of the final grade), due by the beginning of Week 9

Module 4

Final Results and Project Report

Topics

- Combining the literature review with the quantitative analyses
- Building and evaluating the proposed modes from different perspectives
- Listing the limitations of the work
- Revisiting the research questions and work implications to interpret the results

Learning Objectives

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

- Highlight the differences between their results and similar ones in the literature review
- Fully interpret the achieved results of their proposed methodology
- Evaluate the shortcomings of their research and how to improve it

Assessments

- Final Results and Project Report (35% of the final grade), due by the beginning of Week 12

Module 5

Final Presentation

Topics

- Recapping the implemented approach
- Summarizing the research findings and conducted analyses
- Concluding the project and listing future work endeavours

Learning Objectives

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

- Design a project demo to showcase their methods and tools
- Show a thorough understanding of the topic under study
- Provide insight into the optimization of the work
- Show awareness of the current status of their work among other practitioners

Assessments

- Final Presentation (25% of the final grade), due by the beginning of Week 13

Course Textbooks and Materials

Required Textbook(s)

There is no required textbook for this course.

Recommended Readings or Resources

A list of public repositories of data will be provided to conduct your research.

Marking Scheme

Assessment	Course Weight (in %)	Week Due
Project Abstract	5	3
Literature Review, Data Description, and Approach	25	6
Initial Results and the Code	10	9
Final Results and Project Report	35	12
Final Presentation	25	13
Total	**100%**	

- For more information about exams, please see Ryerson University's Senate Policy on Examinations No. 135.

- For more information about grade appeals and reassessments, please see Ryerson University's Senate Policy on Undergraduate Academic Consideration and Appeals no. 134.

Assignment Descriptions

Project Abstract (5% of the final grade)

An abstract is a brief summary of the capstone project. The abstract should contain:

- A brief context about the problem and the theme(s) chosen
- The problem to be solved (e.g., the research questions or the summary of research questions)
- The data to be used
- The techniques (e.g., classification, clustering, text mining, model evaluation etc.) and the tools being proposed to solve the stated problem

The abstract should:

- Include a cover page that includes the student's name, student number, supervisor's name, and date of submission
- Be approximately 500 words (excluding the cover page and references), double-spaced, Times New Roman 12-pt font, using APA writing conventions where appropriate
- Have a references page, using APA style

Literature Review, Data Description, and Approach (25% of the final grade)

Students will conduct research, and review papers and technical articles related to the same problem they have indicated in the abstract. The five basic steps involved in a literature review should be considered:

- Defining research questions
- Introducing the approach
- Searching the literature
- Analyzing materials
- Managing results

It is not preferable to include the source code in the document. Please consider providing a proper link to the project account on the GitHub repository.

Initial Results and the Code (10% of the final grade)

After providing descriptive statistics of the working dataset, applicable attributes, correlation between features, interesting trends (such as outliers and possible reasons), etc., students will think about simple replications/simulations based on their literature review for their dataset. Students can also come up with new approaches, apply these approaches, and document their findings. A set of technical reports are to be submitted, including well-commented and structured source code files.

Final Results and Project Report (35% of the final grade)

After understanding the information being used to develop the proposed machine learning models, students will evaluate these models from three areas: effectiveness, efficiency, and stability.

Students need to statistically compare these models to decide which model outperforms the others. They are required to list the limitations and the implications of their project, and insights on how to improve the work.

The final report should contain:

- A revised version of the research questions
- The main contribution of the work compared to past research
- A link to the GitHub repository where the source code files are checked in
- A more precise description of the applied methodology and the study design
- The conducted analyses, including all activities and their business rules
- A list of all the findings, including a detailed interpretation of the results of the applied techniques
- The shortcomings of the work and concluding remarks on the continuity of the work
- A references page, using APA style

Final Presentation (25% of the final grade)

Students will design a project demo using a Toronto Metropolitan University template to showcase their methods and tools and their understanding of the topic under study. They also need to demonstrate all their project's accomplishments, focusing more on what they have achieved after the literature review submission and how they have answered the proposed research questions. All final presentation sessions will be conducted remotely, according to a schedule announced on the D2L course shell. Presentations will be 10-15 minutes in length followed by a 5-minute question period by two examiners.

The final presentation will be evaluated based on the following criteria:

- Clarity of the presentation
- Providing sufficient information
- Effective use of the presentation time
- Responses to the questions on the topic

Participation Details

Although students need to deliver a set of individual-based reports, they are encouraged to work closely with their designated supervisors. Students are recommended to: i. share their work strategies during their research and analysis, ii. ask questions they may encounter while working on the different stages of

the project, or iii. make changes to their project based on their supervisor feedback. The conversation between a student and a supervisor benefits the students in enhancing their learning and helping practice critical thinking. Discussing content in a live and synchronous environment allows supervisors to address the roadblocks students encounter and provide insightful and constructive feedback.

Etiquette Guidelines

- Treat online forums as academic, public-speaking spaces. Post comments in the same way you would speak in a traditional classroom – politely and respectfully. Forums are a place for discussion and debate about the content you are studying. They are a way of getting to know and interact with your peers and instructor(s) and share your views and ideas
- Respect diversity. There will be multiple perspectives and experiences shared relating to course content and subject matter practice. You may disagree with someone's perspective or have a different one, but positioning any perspective as "right" or "wrong" should be avoided.
- The instructor is the course expert and will address any incorrect information in forums with guidance and support as needed.
- Read and respond to peer postings. If someone comments on your thread or asks a question, monitor and reply.
- Keep criticism constructive and positive. Reference course readings and content to make suggestions or recommendations.
- In online discussions, students are expected to comply with Toronto Metropolitan University's Senate Policy on the Student Code of Non-Academic Conduct No. 61. Inappropriate forum behaviour should be reported to the instructor immediately. Allow the instructor time to respond and take action. Do not engage an inappropriate peer directly.
- Be concise. You, your instructor, and your peers have many posts to read each week. Unless your instructor states otherwise, keep your initial postings and responses brief and meaningful (one to two short paragraphs).
- Your instructor may provide a separate course Q & A forum. This is the ideal place to post general questions about assignments and schedules and to seek clarification on forum issues. Questions or items that are personal in nature, should be communicated with your instructor outside of this forum.

Late Assignments

Late submissions will be penalized with 10% loss a day, up to a maximum of five business days. Exemptions can be made based on conditions stated in the "Missed Term Work" page.

Missed Term Work or Examinations and Course Repeats

Missed Term Work or Examinations

Students are expected to complete all assignments, tests, and exams within the timeframes and by the dates indicated in this Course Outline. Exemption or deferral of an assignment, term test, or final

examination is only permitted for a medical or personal emergency or religious observance (the request must be received within the first two weeks of the course). The instructor must be notified by email **prior to the due date or test/exam date** or as soon as possible after the date, and the appropriate documentation must be submitted. For absence on medical or religious-observance grounds, official forms may be downloaded from the Toronto Metropolitan University website or picked up from The Chang School at Heaslip House, 297 Victoria Street, Main Floor.

Course Repeats

Senate GPA Policy prevents students from taking a course more than three times. For the complete GPA Policy, see Toronto Metropolitan University's Senate Policy on Undergraduate Grading, Promotion, and Academic Standing Policy No. 46.

Plagiarism

The Toronto Metropolitan University Student Code of Academic Conduct defines plagiarism and the sanctions against students who plagiarize. All students are strongly encouraged to go to the Toronto Metropolitan University Academic Integrity Office website and complete the tutorial on plagiarism.

For more information, please consult Toronto Metropolitan University's Senate Policy on Academic Integrity No. 60, 60-1, 60-2 and 61.

Use of the Plagiarism-Detection Service

The work submitted by students in this course will be submitted to Turnitin. Students who do not want their work submitted to this plagiarism-detection service must consult with the instructor to make alternate arrangements by the end of the second module.

Departmental Policies and Course Practices

To learn more about course management expectations, please review Ryerson University's Senate Policy on Course Management No.166.

Accessibility

Per Ryerson University's Senate Policy for Academic Accommodation of Students with Disabilities No. 159, the University will provide academic accommodations for students with disabilities in accordance with the Ontario *Human Rights Code* and the *Accessibility for Ontarians with Disabilities Act*.

If such accommodation is required, please contact Academic Accommodation Support, preferably before the start of the course to allow for time to make any necessary arrangements.

Please Note

Some third party software used in this course may not be accessible to some assistive technology users. Please see the Specific Details on IT Requirements page for more details.

Specific Details on IT Requirements

This checklist will ensure you have the necessary computer and software requirements to successfully complete this course:

- I have access to a reliable computer with the minimum recommended requirements (dual-core 1.6 GHz or faster processor, 200GB of disk, 8GB of RAM) to use the software in this course (e.g. Python).
- I have internet access and adequate data allowance.
- I can access Google Meet and/or Zoom and know how to use it.

Remote Platforms

Each student will be granted access to a Google Cloud Platform (GCP) equipped with RStudio, Python Jupyter Notebook to complete their project. These platforms will be available until the end of the term and will be archived by then. The GCP links will be available and posted on the course shell during the first week of the course.

Please Note

While efforts have been made to make this course accessible to students using assistive technology, the software used in this course is not fully accessible, and there is no comparable software application to date. Software delivered through the virtual desktop will not be accessible for screen reader users.

Student Email

All students in full- and part-time graduate and undergraduate degree programs and all continuing education students are required to activate and maintain their Toronto Metropolitan University online identity at torontomu.ca/accounts in order to regularly access Toronto Metropolitan University's email, RAMSS, the my.ryerson.ca portal and learning system, and other systems by which they will receive official university communications.

Student Support

If you are experiencing technical or administrative issues with your course, help is available from Student Support for Distance Courses via email at distance@ryerson.ca or by phone from Monday to Friday, 9:00 a.m.–5:00 p.m., at (416) 979-5315.