MSc Program in Data Science and Analytics
Major Research Project

MRP Supervision:

The Major Research Project (MRP) may be conducted on various domains/ topics that involves data science and analytics. The MRP examination consists of a written report followed by an oral presentation. The MRP is designed for students to formulate their own themes/problem using open source datasets. The students should be able to identify a few key questions to be tackled using a dataset and employing machine learning methods, visualization techniques, as well as big data tools and algorithms. The students should be clear on the scope or boundaries of their research and analysis. The students will be given a list of public data repositories to choose their data from.

Role of MRP Supervisor

Students will complete their MRP under the guidance of a faculty supervisor. Students are responsible for selecting/ approaching a supervisor who must be a member of YSGS and listed on affiliated faculty list on the program's website.

The MRP supervisor's responsibilities include:

1. Approval Research Proposal: The students should prepare a research proposal and the supervisor must approve/reject within a week. Once the proposal is approved by the supervisor and the second reader, the supervisor or the student should send a copy of the completed and signed MRP Proposal Approval Form to the Program Administrator, latest by the last working day in April.
2. Guiding student's research/ writing and requiring revisions if necessary: Students should expect that they need to produce multiple drafts to ensure an academically strong manuscript and analysis is submitted for evaluation. The student and supervisor should hold several regular meetings, and be in contact via email to track the progress of the work. The supervisor will establish the schedule of contact with the student.
3. Notification of completed draft ready for Poster Presentation: This is the supervisor's decision in agreement with the Second Reader that no further significant revisions to the MRP is required.
4. Establishing the time and location of the MRP Poster Presentation: This will be done in consultation with the student and the Second Reader. This should take place during the first week of September for students who wish to graduate in the fall convocation.
5. Based on the Poster Presentation, advising the student of any revisions required to the MRP: Minor revisions may be required at this stage. Upon the student's completion of these revisions, the supervisor will get the consent of Second Reader, and then will submit MRP Supervisor/Second Reader report signed by supervisor and second reader to the program office.

Finding a MRP Supervisor

Every student will have a faculty Supervisor for their MRP. The student's faculty advisor, course instructors and program director may assist students to identify potential supervisors. Reviewing affiliated faculty profiles from the affiliated faculty page on the program website is the best way to identify potential Supervisors. Full-time students who will be completing the MRP at the end of Summer terms (August) should have a MRP Supervisor confirmed no later than the last working day of April.

MRP Proposal and Milestone Registration

A student will be formally enrolled in the MRP Milestone in RAMMS by the Program Administrator upon receipt of an approved MRP Proposal Approval Form. The form must be signed by the faculty supervisor, the second reader, and the student.

The Second Reader

The Second Reader will be appointed by the Program Director. The Second Reader will be the same person for all MRPs to ensure consistency across the board. The Second Reader, however, is not the co-supervisor. The Supervisor should consult and get concurrence of the Second Reader before approving the research proposal. Once approved, The Supervisor should set major milestones and get the feedback of the Second Reader at those milestones. Upon agreement with the student and the Supervisor, the Second Reader may read earlier drafts. The Second Reader may recommend revisions to the MRP and must agree that it is ready for Poster Presentation before the session is scheduled. Based on the MRP Poster Presentation, both the Second Reader and Supervisor may identify required additional revisions to the MRP before it is approved.

In the rare event that the Supervisor and the Second Reader disagree whether the MRP is ready for Poster Presentation, the Supervisor will involve the Program Director either to resolve the case or hold an additional Poster Presentation with the Program Director.

MRP Poster Presentation

The Poster Presentation is the final stage of the student's work. Students should complete the Poster Presentation before the end of the last term of program registration. Before the Poster Presentation date, the student is responsible for delivering a hard copy of the MRP to the Supervisor and Second Reader. The Poster Presentation is a formal presentation of the MRP. It takes 5 minutes of formal presentation and 5 minutes for questions and answers.

Following the Poster Presentation, the students should expect to be required to make some revisions to the MRP. These will be communicated to the student and be reviewed by the Faculty Supervisor and the Second Reader. The Supervisor will confirm that all final revisions arising from the Poster Presentation have been completed by signing the MRP Supervisor and Second Reader report and forwarding it to the Program Administrator. The report must be received by the Program Administrator before the student's MRP may be accepted to clear the graduation requirements. Students who are unable to clear the requirements to graduate before the published deadline will be enrolled in the program in the following term and are responsible for paying the term fees.

Evaluation

The MRP is graded as pass/ fail. No letter grades are assigned to this course. Assignment of the pass/fail grade in the "MRP Supervisor and Second Reader report" form will be jointly determined by the Supervisor and the Second Reader.

Guidelines for the Major Research Project
MSc Data Science and Analytics
Yeates School of Graduate Studies
Ryerson University

Following are the initial instructions to get you started:

1. Identify the domain, research topic, and research questions that you want to work on.

2. No private datasets are allowed due to the delay in legal process of filing a non-disclosure agreement with the university. Students may work on the projects/ datasets of Data Science Lab by signing a Non-Disclosure and Intellectual Property Rights Agreements with Ryerson University.

3. Here is a list of public repositories of data that you can use for your projects. You need to select one of the following dataset or any other public dataset that match one of the above themes that you selected.

   U.S. Government's Open Data

   Yelp Open Dataset

   U.S. Government's Health Data

   Explore & Download Medicare Provider Data

   City of Toronto Open Data

   Ontario's Open Government Data

   City of New York Open Data

   U.S General Services Administration Open Data

   Government of Canada Open Data

   Statistics Canada Data

   Government of Canada Historical Climate Data

   UCI Machine Learning Repository

   GitHub Archive

   Kaggle Competitions

   MIMIC Open Datasets

   AI Resource Center Open Datasets

   SQL Cookbook to extract data from MIMIC: MIMIC Cookbook

   You might use the beta version of the search engine that Google launches to help scientists find the datasets they need: Google Data Search

4. <u>Note that supervisors will not be coding for you, they will provide you technical and theoretical guidance during the project. You shall be building the product yourself.</u>

5. A data scientist also possesses superior communication skills. You will also need to document this project. We shall provide you the instructions for the documentation later on separately during the DS8005 course. In general, you are building a data product and you will provide an overview of the overall architecture of your product and the results that you get. The project code will be shared using a public GitHub.

([GitHub](#)) repository between you and the instructor. If you would like to use any other repository, then discuss it with the supervisor and the second reader.

6. Late submissions will be penalized with 20% loss of grade for the submitted part of the project. The student must achieve a minimum of B- to get the pass grade.
7. Your project may be used as part of a research paper in future. In that case, you will be one of the co-authors of the paper.

Date	Objective	Deliverable	Marks
April 28, 2024	A. Problem Definition, Data Set and Research Question	A Word or PDF document on the course shell	Mandatory
June 16, 2024	B. Literature Review and Exploratory Data Analysis	A word or PDF document on the course shell	15
July 7, 2024	C. Methodology and Experiments	A word or PDF document on the course shell	25
July 28, 2024	D. Results	A word or PDF document on the course shell, and a Github link	20
August 18, 2024	E. Project Report	A word or PDF document on the course shell	20
TBA	F. Poster Presentations (5+5 minutes)	To be presented in the Poster Live Session	20

A. In the beginning, you need to identify the problem definition, the dataset, and your research questions. You will need to write an abstract of 100 words or less about your project including the problem you are solving, the data set and the techniques and tools you will use. This milestone must be approved both by your supervisor and your second reader in order for you to proceed to the next step.
B. Make a literature review and exploratory data analysis with visualizations.
C. You will need to demonstrate the step-by-step instructions given in Chapter 19 of "Introduction to Machine Learning" by Ethem Alpaydin to meet the objective of methodology and experiments.
D. You will need to provide a Github link including all your codes.
E. You will need to write your report to have the following structure:

a.   Structured Abstract
i.   Background
      ii. Aim
    iii. Methodology
    iv. Results
     v. Conclusions
    b. Introduction
    c. Background and Literature Review
    d. Methodology

e. Results and Discussions

f. Conclusions and Future Work

In addition, your final report will need to contain all front matter formatted according to the YSGS formatting guidelines (please see samples and links provided below), including title page, author's declaration, abstract (coming before structured abstract), any acknowledgments, and table of contents.

F. You will present your project during the poster presentation session to your supervisor and your second reader. The presentation should be 5 minutes long and will have 5 minutes of questions session.

Github

Tutorial: [GitHub Tutorial](#)
Github on Windows: [GitHub on Windows](#)

The responsibility for submitting a MRP in the correct format rests solely with the author.

The formatting guidelines are published online in the Yeates School of Graduate Studies "Thesis and Dissertation Submission Guidelines", starting in <u>Section 5 Formatting</u>:

<u>https://www.torontomu.ca/graduate/student-guide/academic-matters/dissertation-thesis-exams/</u>.

Author's Declaration: <u>Use the wordings provided later in this document</u>.

The wordings provided in the YSGS Thesis & Dissertation Guidelines, section 2.3 (in page 2) may be used for a MRP Author's declaration. However all reference to a 'thesis' in the declaration has to be changed to MRP or Major Research Paper.

Request for Format Checking

As a courtesy the Program Administrator could check the format of MRP and provide feedback <u>if time permits</u>. For students who will complete the degree requirements within the Summer term (May to August), a soft copy of a draft of the MRP or a working in-progress MRP may be sent by email to datascigrad@ryerson.ca to request for comments on the format of the MRP set up by the author of the paper. The Program Administrator will provide feedback based on the YSGS Thesis & Dissertation formatting guidelines. However, the draft copy of the MRP must have all the major components (Front Matter, a draft or a few pages of the Main Body, Back Matter) and two numbering systems properly set up in the paper.

Binding the MRP

If you would like to keep a professionally bound hard copy of your MRP, contact the RSU CopyRite:

> Located at: SCC-B03, Student Centre, 55 Gould Street
> telephone: 416-979-5264
> email: copyrite@rsuonline.ca
> Web URL: www.rsuonline.ca/services

(A sample of the MRP Title Page)

TITLE OF THE MRP IN CAPITAL LETTERS

by

Full name of the author, including past degrees, location, year

(Sample: Jody Terri Smith, BA, AYZ University, yyyy)

A Major Research Paper

presented to Ryerson University

in partial fulfillment of the requirements for the degree of

Master of Science

in the Program of

Data Science and Analytics

Toronto, Ontario, Canada, yyyy

© Author's Name yyyy

(NOTE: yyyy is the year when the student graduates and not the year of submission.

This page is counted as page i but the number i does not show on this page.

Do NOT include this notation in the title page.)

(A sample of the MRP Author's Declaration Page)

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MAJOR RESEARCH PAPER (MRP)

I hereby declare that I am the sole author of this Major Research Paper. This is a true copy of the MRP, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

 I understand that my MRP may be made electronically available to the public.

<Optional: insert typed name...no signature>

(A Sample the MRP Abstract Page)

The title of the MRP and

CENTERED Student's full name

Master of Science yyyy (year of convocation)

Data Science and Analytics

Ryerson University

ABSTRACT

- The content of the master's MRP abstract **must NOT exceed 150 words** .

- The font size must be a minimum of 10-12 points and 10-15 characters per inch.

- The lines of text **must be double-spaced** and left-hand aligned.

- The left-hand margin should be at least 1" and the remaining three margins at least 1" to

  the main text.

For the complete set of the formatting guidelines for the Abstract page, refer to the YSGS Thesis and Dissertation Submission Guidelines, Section 5.7 Abstract:

http://www.ryerson.ca/graduate/currentstudents/academicmatters/dissertations.html

Click the red hot-link in this web page.

Key words:

"Key Words" that identify at a glance what subjects your paper addresses. If you could pick 3 to 5 singles words or BRIEF phrases to tell a reader what your MRP covers, what would they be?