

# RYERSON UNIVERSITY

**Ted Rogers School of Information Technology Management  
And G. Raymond Chang School of Continuing Education**

## **(C)ITM 760 – Big Data Analytics**

### **COURSE OUTLINE FOR 2020-2021**

#### **1.0 PREREQUISITE(S)**

The prerequisite for this course is ITM 618. Students who do not have the prerequisite will be dropped from the course.

#### **2.0 INSTRUCTOR INFORMATION**

- Name:
- Office Phone Number:
- E-mail address:
- Faculty/course web site(s): <https://my.ryerson.ca>
- Office Location & Consultation hours:
  - Your instructor is available for virtual consultation during scheduled consultation hours. Information on the consultation format is provided in the D2L course shell. If you wish to make an appointment, kindly do so via email to ensure the professor is available.
- E-mail Usage & Limits:

Students are expected to monitor and retrieve messages and information sent through D2L and Ryerson email on a frequent and consistent basis. In accordance with the policy on Ryerson student email accounts ([Policy 157](#)), Ryerson requires that any electronic communication by students to Ryerson faculty or staff be sent from their official Ryerson email account. Messages from other accounts may be disregarded.

#### **3.0 CALENDAR COURSE DESCRIPTION**

The objective of this course is to introduce topics in business analytics and data mining with special emphasis on Big Data. Topics may include, but are not limited to, Big Data processing systems, Big Data visualization, data stream mining and large-scale machine learning. Applications will be drawn from various areas such as social network analysis, recommendation systems, and web analytics. Students will gain knowledge on the practical design principles of big data-driven solutions for purposes of business analytics.

#### **4.0 COURSE OBJECTIVES AND LEARNING OUTCOMES**

Learning outcomes describe what students are expected to have learned or achieved; as a result, they usually describe what students will be capable of doing, or what evidence will be provided to substantiate learning.

With rapid advances in computing power and dramatic expansion of data collection and storage capability, nowadays, businesses and organizations have collected vast amounts of data about their business processes. These data are modern-day treasure stores that can be mined to glean insights into a business' products, services and customers. Despite great efforts that have been made in the field of data analytics, there are still many challenges while transcending towards Big Data. How can we discover actionable knowledge from dynamically changing data? How can we effectively combine human and machine intelligence to gain more effective insights from Big Data? This course is a business centric approach to Big Data Analytics where the focus is to train, design and development of techniques and technologies that can be used to answer these challenges. This course will discuss key business analytics techniques, frameworks, and strategies as they are employed to analyze large volumes of data and its application to intelligence discovery in different business area.

#### **COURSE OBJECTIVES**

- Identify a business problem and translate it to an analytics problem and design data-driven solutions.
- Understand the unique challenges of processing and analyzing Big Data at the theoretical and practical level.
- Understand state of the art methods, practices and technologies behind the Big Data processing systems.
- Utilize data mining and machine learning methods to effectively and efficiently process Big Data to support a wide range of queries including business intelligence and data mining.
- Understand the issues involved in building and designing efficient big data systems, and the strategies, data-structures, and technologies used in the implementation of these systems

#### **5.0 TEXTS & OTHER READING MATERIALS**

**Title:** Mining of Massive Datasets, 2nd Edition

**Author:** Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman

**Publisher:** Cambridge University Press

**ISBN:** 978-1107077232

#### **Suggested/Recommended Textbook**

**Title:** Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking

**Author:** Foster Provost, Tom Fawcett

**Publisher:** O'Reilly Media

**ISBN-13:** 978-1449361327

## 6.0 TEACHING METHODS

In Fall 2020 this course will be taught remotely in virtual classrooms. Instruction will take place at scheduled hours, following the approach outlined in D2L Brightspace. You will not be required to attend the Ryerson University campus to complete this course.

[Click here to enter text.](#)

## 7.0 EVALUATION, ASSESSMENT AND FEEDBACK

The grade for this course is composed of the mark received for each of the following components:

Evaluation Component	Percentage of the Final Grade
Assignments	20%
Course Project	25%
Midterm Examination	20%
Final Examination	35%
<b>Total</b>	<b>100%</b>

**NOTE:** Students must achieve a course grade of at least 50% to pass this course.

- ❖ All assignments submitted for grading will be handed back before Fall and Winter of each semester before the last day to drop a course in [good academic standing](#).
- ❖ At least **20%** of student's grade based on individual work will be returned to students prior to the last date to drop a course without academic penalty.

### Citation Format for Essays and Term Papers

All essay assignments, term paper and other written works must adhere with APA citation format. Technical errors (spelling, punctuation, proofing, grammar, format, and citations) and/or inappropriate levels of language or composition will result in marks being deducted. You are encouraged to obtain assistance from the Writing Centre ([www.ryerson.ca/writingcentre](http://www.ryerson.ca/writingcentre)) for help with your written communications as needed.

You can find APA guidelines and academic referencing from the following online resources:

[Student Learning Support > Online Resources > Writing Support Resources](#)

- [APA Basic Style Guide](#)

[Ryerson Library Citations and Style Guides](#)

- [APA Style](#)

## 8.0 PLAGIARISM DETECTION

Turnitin.com is a plagiarism prevention and detection service to which Ryerson subscribes. It is a tool to assist instructors in determining the similarity between students' work and the work of other students who have submitted papers to the site (at any university), internet sources, and a wide

range of books, journals and other publications. While it does not contain all possible sources, it gives instructors some assurance that students' work is their own. No decisions are made by the service; it generates an "originality report," which instructors must evaluate to judge if something is plagiarized.

Students agree by taking this course that their written work will be subject to submission for textual similarity review to Turnitin.com. Instructors can opt to have student's papers included in the Turnitin.com database or not. Use of the Turnitin.com service is subject to the terms-of-use agreement posted on the Turnitin.com website. Students who do not want their work submitted to this plagiarism detection service must, by the end of the second week of class, consult with their instructor to make alternate arrangements.

Even when an instructor has not indicated that a plagiarism detection service will be used, or when a student has opted out of the plagiarism detection service, if the instructor has reason to suspect that an individual piece of work has been plagiarized, the instructor is permitted to submit that work in a non-identifying way to any plagiarism detection service.

## 9.0 TOPICS – SEQUENCE & SCHEDULE

Session	Topic	Learning Outcomes	Readings	
1	Introduction to Big Data Analytics	•	Get familiarized yourself with data analytics process, Big Data analytics terminology, concepts and challenges	Chapter 1
Lecture notes				
2	MapReduce and the New Software Stack	•	Understand how distributed file systems and MapReduce work	
•	Explain major big data environments such as Apache Hadoop and Apache spark			
•	Describe main challenges in Big Data environments	Chapter 2		
Lecture notes				
3	Mining Data Streams	•	Understand concepts of data streams, data stream mining and main challenges	
•	Explain how to do sampling over data streams	Chapter 3		
Lecture notes				

4	Mining Data Streams	•	Get familiarized with data stream processing models such as sliding window and decaying model	
•	Understand how to filter data streams	Chapter 3		
Lecture notes				
5	Link Analysis	•	Understand how first generation of Web search engines were working	
•	Describe popular link analysis methods such as PageRank	Chapter 5		

### 10.0 VARIATIONS WITHIN A COURSE

All sections of a course (Day and CE sections) will follow the same course outline and will use the same course delivery methods, methods of evaluation, and grading schemes. Any deviations will be posted on D2L Brightspace once approved by the course coordinator.

### 11.0 OTHER COURSE, DEPARTMENTAL, AND UNIVERSITY POLICIES

For more information regarding course management and departmental policies, please consult the [Course Outline Appendix](#) which is posted on the [Ted Rogers School of Information Technology Management website](#).

**NOTE:** Students must adhere to all relevant university policies found in their online course shell in D2L and /or on the following URL: [senate-course-outline-policies](#).

The appendix covers the following topics:

- Attendance & Class Participation
- Email Account
- Request for Academic Consideration
- Examinations & Tests
- Late Assignments
- Standard of Written Work
- Academic Grading Policy
- Academic Integrity
- Student Rights

## Important Resources Available at Ryerson

- [Academic Accommodation Support](#): Ryerson University acknowledges that students have diverse learning styles and a variety of academic needs. If you have a diagnosed disability that impacts your academic experience, connect with Academic Accommodation Support (AAS). Visit the [AAS website](#) or contact [aasadmin@ryerson.ca](mailto:aasadmin@ryerson.ca) for more information. Note: All communication with AAS is voluntary and confidential, and will not appear on your transcript.
- [The Library](#) provides research workshops and individual assistance. If the University is open, there is a Research Help desk on the second floor of the library, or go to [Workshops](#).
- [Student Learning Support](#) offers group-based and individual help with writing, math, study skills, and transition support, as well as [resources and checklists to support students as online learners](#).
- You can submit an [Academic Consideration Request](#) when an extenuating circumstance has occurred that has significantly impacted your ability to fulfill an academic requirement.
- [Ryerson COVID-19 Information and Updates for Students](#) summarizes the variety of resources available to students during the pandemic.
- Familiarize yourself with the tools you will need to use for remote learning. The [Continuity of Learning Guide](#) for students includes guides to completing quizzes or exams in D2L or Respondus, using D2L Brightspace, joining online meetings or lectures, and collaborating with the Google Suite.