CISC 451/839: Topics in Data Analytics
(Fall 2019)

OnQ: TBD

Course Description

In the Big Data era, we all, in one form or another, participate in generating data. Increasingly, organizations' success has become dependent on how quickly and efficiently they can turn these petabytes of data into actionable insights. Data Analytics is the art and science of extracting such insights. However, it requires experimenting-with and integrating different analytics techniques, while handling the Big Data high arrival velocity and large volumes.

This course provides a full overview of the Data Analytics pipeline. The aim of this course is to teach practice rather than theory. The focus will be on using existing libraries and tools to produce solutions. The course is intended to be a hands-on course. The course covers data wrangling techniques, such as dealing with corrupt data, anomalies, missing values, high dimensionality, irregularities, and other issues that a data scientist truly need on a daily basis. You will learn applied predictive modeling methods, as well as how to explore and visualize data, and how to use and understand common machine learning algorithms to come to sound conclusions about your data, despite the real-world challenges. Later in the course, Big Data Analytics techniques are introduced and you will learn how to handle data of large volumes. Finally, Deep Learning is introduced and you will learn how to extract insights from images/videos and text/speech.

Disclaimer: The course will teach you how to use existing machine learning methods, but will not teach how to implement these algorithms from scratch.

CO/Prequisites

This course is designed for those who are analytically minded and are familiar with basic statistics and programming or scripting. This course is to be taken with or subsequent to CISC 121 (programming), CISC 332*/CMPE 332 (databases) and (STAT 269 or CISC/CMPE 333 (Data mining)), CISC 351 or comparable courses.

Learning Outcomes

After finishing this course, you should be able to:

- Clean and prepare your data for analysis.
- Perform basic visualization of your data.
- Model your data using supervised and unsupervised learning.
- Work with Big Data.
- Analyze images/videos and text/speech using Deep Learning.

Programming Languages and Platforms:

We will be exploring the following Analytics tools:

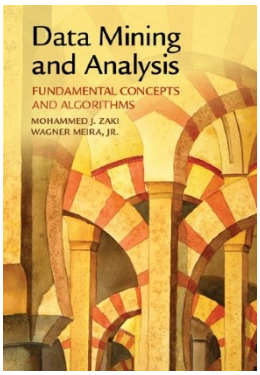
- **Rapid Miner** [https://rapidminer.com/]
- **Weka** [http://www.cs.waikato.ac.nz/ml/weka/]

- **Python** [https://www.python.org/
- **R** [https://www.r-project.org/]

- **Tesnorflow [**https://www.tensorflow.org/**]**
- **Scikit -learn [http://scikit-learn.org/stable/]**
-
- **Hadoop Ecosystem[**http://hadoop.apache.org/**]**
- **Spark Ecosystem[**https://spark.apache.org/**]**

Resources:

We will be using multiple resources in this course including some research papers that you can download online. However, if you like having a text book to study from, I recommend:



Mohammed J. Zaki, Wagner Meira, Jr., *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, May 2014. ISBN: 9780521766333.

A free copy for personal use is available here:
http://www.dataminingbook.info/pmwiki.php/Main/BookDownload

Course Schedule

Week	Topic	Required Readings	Deliverables
1	Introduction	Chapter 1	
2	Data Understanding	Chapter 2 and 3	Online quiz 1
3	Visualization		
4	Data Preparation	Chapter 7	Online quiz 2
5	Classification	Chapters 18, 19, 21 and 22	Competition 1 (Data Preparation) submission
6	Regression		Online quiz 3
7	Clustering	Chapters 13 and 17	Competition 2 (Supervised) submission
8	Big Data (Processing & Storage)		Online quiz 4
9	Deep Learning	Wang, Haohan, Bhiksha Raj, and Eric P. Xing. "On the Origin of Deep Learning." arXiv preprint arXiv:1702.07800 (2017).	Competition 3 (Unsupervised) submission
10	Deep Learning		Online quiz 5
11	Deep Learning		Competition 4 (Big Data Mining) submission
12	Revision & The Analytics Showcase		Online quiz 6 Poster and project presentation for the Analytics Showcase .

Marking Scheme

10 %	Bi-weekly Online Quizzes on OnQ	Six 2% each and only top 5 scores count
30 %	Analytics Showcase Project & Poster	Two deliverables: • Poster [10%] • Code [20%]
20 %	Analytics Competitions	Four competitions, 5% each.
40 %	Final Exam [Must pass]	
5%	Plus possible bonus points	

- **Late Policy:** A point will be deducted for each day after the deadline.
- More details can be found on the course OnQ page.

GRADING METHOD

Grade	NumericalCourse Average (Range)
A+	90-100
A	85-89
A-	80-84
B+	77-79
B	73-76
B-	70-72
C+	67-69
C	63-66
C-	60-62
D+	57-59
D	53-56
D-	50-52
F	49 and below

Course Policies

- Requests for additional work to boost an unsatisfying grade will be denied.
- No retakes for online quizzes.

- Any question having to do with grades is to be directed to the instructor, and NOT to a TA.
- Marks will be reported on OnQ, as they become available.

GENERAL INFORMATION

Weekly Schedule

Day	From	To	Location
Monday	4:30PM	5:30PM	HUMPHREY AUD
Wednesday	3:30PM	4:30PM	HUMPHREY AUD
Friday	2:30PM	3:30PM	HUMPHREY AUD

Instructor and Teaching Assistants

	Name	Email Address	Office	Office Hours
Instructor	Hazem Abbas	Hazem.abbas@queensu.ca	GW623	TBD
TA	Alireza Sedghi	alireza.sedghi@queensu.ca		

LOCATION AND TIMING OF FINAL EXAMINATIONS

As noted in Academic Regulation 8.2.1, "the final examination in any class offered in a term or session (including Summer Term) must be written on the campus on which it was taken, at the end of the appropriate term or session at the time scheduled by the Examinations Office." The exam period is listed in the key dates prior to the start of the academic year in the Faculty of Arts and Science Academic Calendar and on the Office of the University Registrar's webpage. A detailed exam schedule for the Fall Term is posted before the Thanksgiving holiday; for the Winter Term it is posted the Friday before Reading Week, and for the Summer Term the window of dates is noted on the Arts and Science Online syllabus prior to the start of the course. Students should delay finalizing any travel plans until after the examination schedule has been posted. Exams will not be moved or deferred to accommodate employment, travel /holiday plans or flight reservations.

INSTITUTIONAL POLICIES

Academic integrity is constituted by the five core fundamental values of honesty, trust, fairness, respect and responsibility (see www.academicintegrity.org). These values are central to the building, nurturing and sustaining of an academic community in which all members of the community will thrive. Adherence to the values expressed through academic integrity forms a foundation for the "freedom of inquiry and exchange of ideas" essential to the intellectual life of the University (see the Senate/Report on Principles and Priorities)

Students are responsible for familiarizing themselves with the regulations concerning academic integrity and for ensuring that their submitted work conforms to the principles of academic integrity. Information on academic integrity is available in the Arts and Science Calendar (see Academic Regulation 1), on the Arts and Science website and from the instructor of this course.

Departures from academic integrity **include plagiarism, use of unauthorized materials, facilitation, forgery and falsification**, and are antithetical to the development of an academic community at Queen's. Given the seriousness of these matters, **actions which contravene the regulation on academic integrity** carry sanctions that can **range from a warning or the loss of grades on an assignment to the failure of a course to a requirement to withdraw from the university**.

Turnitin Statement: Queen's University has partnered with the third-party application Turnitin to help maintain our standards of excellence in academic integrity. Turnitin is a suite of tools that provide instructors with information about the authenticity of submitted work and facilitates the process of grading. Submitted files are compared against an extensive database of content, and Turnitin produces a similarity report and a similarity score for each assignment. A similarity score is the percentage of a document that is similar to content held within the database. Turnitin does not determine if an instance of plagiarism has occurred. Instead, it gives instructors the information they need to determine the authenticity of work as a part of a larger process.

ACCESSIBILITY

Queen's University is committed to ensuring that information, resources, buildings and services are as accessible as possible. While ensuring accessibility and integration of students with disabilities is everyone's responsibility at Queen's, services for students with disabilities are provided by the Queen's Student Accessibility Services (QSAS). For further information, please visit the QSAS Office Web site: http://www.queensu.ca/studentwellness/accessibility-services

If for any reason you **can't take online tests [medical reasons]**, please come talk to me or send me an email and we will figure something out.

If you need **special accommodations for your final exam**, you need to discuss that with the Queen's Student Accessibility Services (QSAS) and they will help you to arrange it and the **deadline for that is mid-semester**. If you don't bother to make the arrangements by the deadline you'll have to write in the regular exam hall.

If you have a **sudden last-minute need for accommodations (such as a broken arm)** then talk with the Queen's Student Accessibility Services (QSAS) and if they write me a letter confirming that you need accommodations they can't provide I will help you out.