

GGE 5405/GEE 6505 – INTRODUCTION TO BIG DATA & DATA SCIENCE COURSE

SYLLABUS

Instructor:	Nasrin Eshraghi Ivani	Class Day(s):	Mondays and Wednesdays
Email:	nasrin.eshraghi@unb.ca	Time:	9:30am – 10:20am
Office Phone:	1 5064526116	Class Location:	Head Hall 211
Office Location:	Head Hall Room E51C	Office Hours:	Via Appointment

Course Description

This course aims to introduce students to the theoretical and practical fundamentals of Data Science and Big Data in order to encourage critical thinking, problem-solving and computational reasoning. The course is data project-oriented. The in-class portion of the big data project is meant to be collaborative and a time for the instructor to work closely with groups to understand the learning objectives and help them work through any logistics that may be slowing them down. Lectures will introduce a data analytics workflow that will be used in the data project, including data access, data storage, data wrangling, data analytics and data visualization and storytelling. In addition, students will learn new technologies and tools in big data analytics. In this course, students will learn and gain practical experience with big data tools and technologies.

Library information

www.lib.unb.ca

UNB Libraries provides access to a vast collection of online and print resources. Use Research by Subject on the library website to find the best resources for this course. Research help is available by phone, e-mail, chat, and in-person. The libraries offer quiet and group study spaces. Book a Group Study Room online at http://www.lib.unb.ca/services/group_study.php

Online Materials

Online course materials can be found in Desire2Learn (Brightspace), UNB's online Learning Management System. You can access it through the MyUNB portal for a single login to all UNB services (<https://my.unb.ca/group/mycampus/home>) or directly at <https://lms.unb.ca/>.

Further Reading

Godsey, B. (2017). *Think like a data scientist: Tackle the data science process step-by-step*. Manning Press.

Granville, V. (2014). *Developing Analytical talent: Becoming a data scientist*. Wiley Press.

Grus, J. (2015). *Data Science from scratch: First principles with Python*, O'Reilly Media.

Hand, D.J. ((2016). *Measurement: A very short introduction*. Oxford Press.

James G., Witten D., Hastiew T. & Tibshirani R. (2013). An introduction to statistical learning with applications in R. Springer.

Nielsen, L. (2015). A simple introduction to Data Science: Book 2. New Street Communications, LLC
Yu S., Lin X., Misic J., & Shen X. (2015). Networking for Big Data. CRC Press.

Wes, M. (2012). Python for data analysis. O'Reilly Media.

Wessler, M. (2018). Big Data Management for Dummies, 2nd Informatica Special Edition, A Wiley Brand. Available in D2L.

Boschetti, A., & Massaron, L. (2015). Python data science essentials: become an efficient data science practitioner by thoroughly understanding the key concepts of python (Ser. Community experience distilled). Packt Publishing. Retrieved October 12, 2021, from INSERT-MISSING-URL.

Course Outcomes

The course consists of ten main modules. They are:

1. Introduction to Data Science and Big Data
2. Data Access
3. Data Storage and retrieval
4. Data Wrangling
5. Data Visualization
6. Storytelling with data
7. Data Analytics
8. Data Streaming
9. Big Data Pipeline and Tools
10. Big data Analytics

Upon completion of this course, you will be able to:

- understand what Data Science is, and learn where big data originate and what are the key challenges;

- learn the Big Data pipeline and the most used techniques in data engineering.
- focus on the big data questions and answers— the value is in the answers, not the data.
- develop the ability and knowledge to install the appropriate software yourself.
- develop big data analytical workflows and programming models for batch, interactive, and streaming applications.
- comprehend spatial computing and algorithms for big data analytics.
- describe the principles of tools for data visualization and data-oriented user interfaces.
- understand the computational complexity in data science when it applies to linear and non-linear models.
- develop critical thinking skills for becoming a data scientist.

Students' competency levels on these outcomes may vary. Outcome achievement requires the meeting of all course expectations, including honoring all course policies, regular class attendance, and completion of all assigned work in good faith and on time.

Grading Scale

90-100 = A+	85-89 = A	80-84 = A-	77-79 = B+	73-76 = B	70-72 = B-	
67-69 = C+	63-66 = C	60-62 = C-	57-59 = D+	53-56 = D	50-52 = D-	0-49 = F = 0

Assignments

The course has seven assignments which will provide the student with hands-on experience on data science tools and technologies. They will provide the learning experience with big data challenges, and help the student to develop critical thinking skills.

Final Project

The course has a research project that aims to enrich your understanding of how to set up a big data project that requires understanding the value of big data as well as big data technologies that can be used to store and analyze data. Based on your literature review, you will select a real-world use case to introduce your problem domain and choose the optimal big data tools for given requirements, the relevant technologies for handling big data, and the type of data analytics and expected results.

REAL-WORLD USE CASES

- Data Science for smart cities
- Data Science contributions to healthcare
- Data Science for finance and crime prediction
- Data Science for retail
- Data Science for Social Good
- Data Science for tourism
- Data Science for transportation
- Data Science for manufacturing

TASKS

- Literature review on a selected real-world use case.
- Define an interesting question about your real-world use case that you might be able to answer using big data.
- Select the type of data you need to analyze in order to answer your question.
- Develop a big data analytical workflow that will consist of a description of main requirements and expected outcomes for each one of the steps including data access, data measurement, data wrangling, data analytics, and data visualization.

OUTCOMES

- Project paper (5-10 pages, **only for GGE6505**)
- Presentation where you will outline your big data project (20 min).

Course Marking Scheme

Assignment	Weight (GGE6505)	Weight (GGE5405)
Assignments and Hands-on	40	60
Seminar participation	10	10
Big Data Project Report	30	-
Big Data Project Presentation	20	30
Total:	100	100

Services for Students with Disabilities

If you are a student with a disability of any type (physical, mental, learning, medical, chronic health, sensory; visible or invisible) you are strongly encouraged to register with the UNBF Student Accessibility Centre (SAC) (<http://www.unb.ca/fredericton/studentservices/academics/accessibility/>) so that you may receive appropriate services and accommodations. Once you are registered with SAC, the instructor will be notified via the UNBF SAC Accommodation Letter of your specific accommodations. If you would like to discuss your particular needs with the instructor, please book a time for a confidential appointment.

Writing and Study Skills Support

UNB's Student Services provides many coaching and mentoring services to assist with writing papers, effective study methods, and other skills development related to student success:

<http://www.unb.ca/fredericton/studentservices/academics/writing-centre/index.html>

Math Skills Support

UNB's Math Learning Centre offers math help drop-in times and opportunity to book appointments:

<http://www.math.unb.ca/~mathhelp/>

Academic Offences

Academic offences include, but are not limited to, the following:

Plagiarism

Plagiarism includes:

1. Quoting verbatim or almost verbatim from any source, regardless of format, without acknowledgement;
2. Adopting someone else's line of thought, argument, arrangement, or supporting evidence (such as statistics, bibliographies, etc.) without indicating such dependence;
3. Submitting someone else's work, in whatever form (essay, film, workbook, artwork, computer materials, etc.) without acknowledgement;
4. Knowingly representing as one's own work any idea of another.

NOTE: In courses that include group work, a penalty may be imposed on all members of the group unless an act of plagiarism is identified clearly with an individual student or students.

Examples of other academic offences include cheating on exams, tests, assignments or reports; impersonating somebody at a test or exam; obtaining an exam, test or other course materials through theft, collusion, purchase or another improper manner, submitting course work that is identical or substantially similar to work that has been submitted for another course; and more as set out in the academic regulations found in the Undergraduate Calendar. Penalties for plagiarism and other academic offences range from a minimum of F (zero) in the assignment, exam or test to a maximum of suspension or

expulsion from the University, plus a notation of the academic offence on the student's transcript. For more information, please see the Undergraduate Calendar, University-Wide Academic Regulations, Regulation VIII.A, or visit: <http://go.unb.ca/tlsPb0XX5>. It is the student's responsibility to know the regulations.