College of Liberal Arts and Sciences
Department of Computer Science & Physics
CSC 410, Data Engineering
CRN: 21912, Section: I1, Spring 2020, Credits: 3.0
Building: SCI 220, Day: MW, Time: 02:50 pm – 04:20 pm

Instructor Information

Instructor: Dr. Md Liakat Ali (Dr./Prof. Ali)
Office: Science Hall, Room 204 D
Office Hours: MON, TUE, THU 12-1:00 pm, or by appointment
E-mail: mdali@rider.edu

Course Description: This course serves as an introduction to the interdisciplinary and emerging fields of data engineering and data science. Students learn to combine tools and techniques from computer science, statistics, data visualization and the social sciences to solve problems using data. Central themes include: the data engineering and data science processes; tools for working with both big and small datasets, statistical modeling, and machine learning. Specific topics and tools include: data wrangling and munging, machine learning algorithms, statistical models, data visualization, data pipelines, ethics, Hadoop, Spark, R, Python, and MapReduce. **Prerequisites and Restrictions:** CSC 410 is a required course for CS majors. The prerequisites for this course is CSC 230 Probability for Computer Science.

Textbook: There is no required specific text book for the class, but it is highly recommended that you should have one of the following books:

1. Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, by EMC Education Services, Wiley and Sons. (2015).
2. Introducing Data Science: Big Data, Machine Learning, and More, using Python Tools, by Davy Cielen, Arno D. B. Meysman, and Mohamed Ali, Manning Publications. (2016). Available in Online: https://www.manning.com/books/introducing-data-science?query=Introducing%20Data%20Science

Course objectives: Upon completion of this course, the student should be able to:

- Explain the difference between data engineering, data science and data analytics.
- Describe the various types of data: structured, unstructured, natural language, machine-generated, graph-based or networked, audio, image, video, and streaming.
- Recognize problems solvable with machine learning and data science algorithms.
- Explain and use data visualization techniques.
- Discuss about classification, clustering, text mining, and information extraction
- Analyze the distributing data storage and processing with frameworks, Hadoop, MapReduce, and NoSQL ("non SQL" or "non-relational") databases and data modeling.
- Discuss and use of Python and Python data engineering libraries, Scikit-learn and StatsModels.

Requirements and expectations:

- Canvas: Canvas will be used to provide and submit all homework assignments, quizzes and grades. Make sure that you check your score in Canvas regularly. Any questions regarding Canvas points, you have to ask/email me **before Final Exam**.

- Software: Latest version of R and RStudio Desktop: available at-https://cran.rstudio.com/, https://rstudio.com/products/rstudio/download/. Latest version of Python available at https://www.python.org/downloads/ Latest version of PyCharm Python IDE available at https://www.jetbrains.com/pycharm/
- scikit-learn Machine Learning in Python: https://scikit-learn.org/stable/
- Commitment to attend class and log-on periodically to Canvas to check for announcements, check calendar, post, read and respond to messages in the discussion area.
- During all classes you are expected to take substantive notes and to develop a list of terms/concepts as instructed by the instructor.
- For ALL assignments you will be expected to submit it to me by the due date, no assignments will be accepted after the due date. All work must be your own; the rules for Academic Honesty are not to be violated (see Rider University Student Handbook). Anyone that violates the rules is subject to get a zero for the assignment/test and an F for the course. In addition you must adhere to the Rider academic policy, which can be viewed here: http://catalog.rider.edu/policies/code-academic-integrity/

Classroom Decorum:

- Be on time! Class will begin promptly at the scheduled time. Allow yourself enough time to get to class, ready to learn, before the period begins. Do your best to remain in the room during the period. Exiting and entering during the period breaks the concentration of your fellow students, and makes it hard for you to get the full value of the class.
- Turn off all cell phones, no earbuds, no texting, and anything else that would cause a distraction to yourself or others around you. If there's an emergency situation, you have to inform your instructor.
- Participate in class discussions; ask questions, no eating or drinking in class. Be aware of body language; it can speak volumes! No sleeping in class
- Students are permitted to use computers/laptops during class for note-taking and other class-related work only. Those using computers/laptops during class for purposes not related to the class (like e-mailing, instant messaging, game playing or internet surfing) will be asked to leave the classroom for the remainder of the class period.

Graded Course Activities: Visit Canvas for details about each assignment, class work, project and exam listed below. All class works and assignments for this course will be submitted electronically through Canvas unless otherwise instructed. Points you receive for graded activities will be posted to the Canvas.

Participation: ATTENDANCE IS MANDATORY. Class participation is worth 05% of your final grade. Class participation is critical in this course. Much of the work is additive/cumulative. Consequently, frequent absences will adversely affect your grade. You will also lose credits for class-work points if you miss classes. If you have perfect attendance, your Professor will consider boosting your grade if it is borderline. If you miss a class, it will be your responsibility to determine what you missed and to make up the work. In addition, you must adhere to the Rider attendance policy.

Assignment: There are six assignments in this course. Each assignment should be submitted on the Canvas website assignment page. Completed assignments must be submitted electronically via canvas no later than 11:59 PM on the due date listed in the syllabus. No homework will be accepted after the due date. **Do not email me your assignment, as it will not be accepted**.

Exam: One midterm and one final exam will be given in the semester session. The exams will test assigned readings and material discussed in class. You are not allowed to copy others work. The final exam will **not** be comprehensive in nature and will cover the material after the second test. All exams will be **'closed-book'** format. Students who are absent during a class period when a test is given, will receive a score of zero unless previous arrangements are made or a valid written excuse is presented.

Class-work: There are six class-works in this course. If you miss a class, you will miss class-work points. There will be no makeup. Proofs are needed for exceptions or true emergencies.

Project: There is a team project with maximum two persons in each team. The project should involve analytics on real world data of significant size and must be approved by the instructor. Visit canvas to read more about project guidelines.

Letter Grade Assignment: Final grades assigned for this course will be based on the percentage of total points earned and are assigned as follows:

Grading Criteria	
Points	**Description**
18	Assignment
09	Classwork
20	Project
18	Mid Term
30	Final Exam
05	Class Participation
100	Total Points Possible

Letter Grade	**Percentage**	**Quality Points**
A	93.0-100%	4.0
A-	90.0-92.9%	3.7
B+	87.0-89.0%	3.3
B	83.0-86.9%	3
B-	80.0-82.9%	2.7
C+	77.0-79.9%	2.3
C	73.0-76.9%	2
C-	70.0-72.9%	1.7
D	60.0-69.9%	1
F	0-59.9%	0

Academic Honesty: In accordance with the Rider University rules of Academic Honesty, students must complete their individual work on their own. Academic dishonesty includes any unauthorized collaboration or misrepresentation in the submission of academic work. In all written work, whether in class or out of class, the student's name on the work is considered to be a statement that the work is his or hers alone, except as otherwise indicated. Students are expected to provide proper citations for the statements and ideas of others whether submitted word for word or paraphrased. Please note that all ideas, phrases, sentences taken from the Internet and other sources and incorporated into your own writing must be cited to avoid plagiarism. Disregard for this rule will result not only in failure, but also in a letter being placed in your permanent file. Please note that incorporating material from class-discussions, lectures and our suggestions for improvement does NOT constitute academic dishonesty in this class.

Accommodations for Students with Disabilities: If you have a disability and believe you will need academic accommodations in this course, please make an appointment for an Intake Interview with Services for Students with Disabilities in the Vona Academic Annex, Room 8. The phone number is (609) 895-5492 and the email is serv4dstu@rider.edu. To learn more, please visit:https://www.rider.edu/academics/academic-support-services/student-accessibility-support-services

Academic Issue: If at any time a student has an academic problem in any course with a grade

or any other issue, the student's first course of action to resolve the matter should be to make an appointment with the instructor to discuss the issue.

Content Calendar

Week	Date	Topic	Class Work	Assignment
1	01/27/2020	Introduction		
	01/29/2020	Introduction to Big Data Analytics		1
2	02/03/2020	Data Analytics Lifecycle		Due: 02/09/20
	02/05/2020			
3	02/10/2020	Review of Basic Data Analytic Methods		
	02/12/2020	Using R **Project proposal presentation**		
4	02/17/2020	Advanced Analytical Theory and Methods:		2
	02/19/2020	Clustering		Due: 02/25/20
5	02/24/2020		1	
	02/26/2020	Advanced Analytical Theory and Methods:		3
6	03/02/2020	Association Rules		Due: 03/08/20
	03/04/2020	Mid Term review	2	
7	03/09/2020	**Mid Term Exam**		
	03/11/2020	**Mid Project Presentation**		
8	03/16/2020	**No Class (Spring Recess)**		
	03/18/2020			
9	03/23/2020	Advanced Analytical Theory and Methods:		4
	03/25/2020	Regression		Due: 04/01/20
10	03/30/2020		3	
	04/01/2020	Advanced Analytical Theory and Methods:		5
11	04/06/2020	Classification (Decision Trees)	4	Due: 04/12/20
	04/08/2020	Advanced Analytical Theory and Methods:		6
12	04/13/2020	Classification (Naïve Bayes)	5	Due: 04/19/20
	04/15/2020	Advanced Analytical Theory and Methods:		
13	04/20/2020	Text Analysis	6	
	04/22/2020	Advanced Analytics-Technology and Tools:		
14	04/27/2020	MapReduce and Hadoop/ Students' Presentation		
	04/29/2020	**Final Exam**		
15	**Final Project Presentation: May 7, 2020 Time: 3:30-5:30 PM**			

Important Note: This syllabus, along with course assignments and due dates, are subject to change. It is the student's responsibility to check Canvas for corrections or updates to the syllabus. Any changes will be clearly noted in course announcement or through email.