CS 842 Introduction to Data Science - Winter 2021

University of Regina, Department of Computer Science
November 13, 2020

Schedule

- **Lecture:** Blended - Asynchronous Online Video Lectures + QnA sessions via Zoom during class schedule
- **Schedule:** TR 11:30 - 12:45
  Jan 12, 2021 to April 15th, 2021. No classes on Feb 16 and 18 (Winter break).
- **Project Milestones:**
  February 10th, 2021 (Project Proposal Presentation Due)
  April 10th, 2021 (Final Project Presentation Due)

Instructor

Dr. Alireza Manashty

- **Location:** CW 308.25 (online: [https://zoom.us/my/manashty](https://zoom.us/my/manashty))
- **Hours:** Flexible as needed with appointment request via UR Courses or Email.
- **Email:** alireza.manashty@uregina.ca

Course Description

Introduction to Data Science provides a broad overview of the data science process lifecycle which includes data discovery, data preparation, model planning, machine learning model implementation and evaluation, visualization, and delivery. The course provides hands-on data science experience via a real-world project.

Textbooks

There are several books and many online resources available for this topic. The following books are considered to be the main sources of reference. More online resources will be provided in UR Courses.

**Mandatory, Primary Textbook**


EMC Education Services, Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, 2015, 9781118876138
Evaluation

<table>
<thead>
<tr>
<th>Assignments</th>
<th>Project</th>
<th>Final Exam</th>
<th>Instructor’s discretion</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>50%</td>
<td>30%</td>
<td>±5%</td>
</tr>
</tbody>
</table>

Your final grade will be a representative of the University of Regina grading scheme located at [https://www.uregina.ca/gradstudies/current-students/grad-calendar/grading-system.html](https://www.uregina.ca/gradstudies/current-students/grad-calendar/grading-system.html)

Tentative Topics/Week:

- Introduction to Data Science and Big Data Analytics / Session 1
- Data Science Process Lifecycle / Session 2
- Business Problem Discovery / Session 2
- Data Acquisition, Preparation, and Preprocessing / Session 3 and 4
- Model Planning / Session 6
  - Statistical and Machine Learning Models / Session 6, 7
  - Feature Engineering / Session 8
  - Model Implementation and Evaluation / Session 9
  - Introduction to Deep Learning models / Session 10
- Visualization and Presentation / Session 11
- Project Delivery and Operationalization in Cloud / Session 12
- Project Presentations and Discussions / Session 13

Assignments

Assignments must be completed individually. Instructions for each assignment along with the method of submission will be provided in UR Courses. All assignments must be delivered on the due date posted for each assignment. There will be five assignments: Python (2%), Quizzes (6%), and 3 x 4%. You will need to have access to a computer running JupyterLab or similar programming notebooks using Python 3.X programming language. You can install the JupyterLab from here: [https://jupyter.org/install](https://jupyter.org/install)
Project

A real-world data science project should be proposed, implemented, and presented by the students. Details and rubrics will be provided in UR Courses. The students are expected to be familiar with or learn the Python programming language. Project proposal and final project include both a presentation and a report. Projects can be completed individually or in groups of two or three students. Your final project grade will be adjusted by +10%, +5%, and 0% depending on the size of your group (one, two, or three), respectively. If the adjusted project grade goes over 100%, you may or may not receive a bonus mark (as per instructor’s discretion). Based on the FGSR grading (see above), a publishable work will receive 95%-100% on the project. Due to the size and the complexity of the projects, a group size of two is recommended.

Final Exam

The exam will be an online exam in UR Courses during the scheduled exam period on April 20th, 2021. This course will use the Proctortrack remote proctoring platform for verification of student identity and monitoring of class exams. As a result, your personal information will be securely and temporarily collected and stored under the legal authority of The Local Authority Freedom of Information and Protection of Privacy Act. This personal information will include your first and last name, institution name, student number, image, as well as recordings of you and your computer screen during a proctored exam. Students are advised to complete the Student Onboarding process early in the term to allow as much time as possible to resolve any issues that may arise. Further information can be found here: https://www.uregina.ca/remote-learning/index.html#proctoring

Important Note: You must pass the final examination in order to pass the class.

Bonus presentations (if any):

Bonus presentations (should see/email me before February 5th), or exceptional performance in assignments, project, or final exam could earn up to 2%.

Plagiarism

Please refer to University of Regina policy for plagiarism and academic integrity in here. https://www.uregina.ca/gradstudies/current-students/grad-calendar/policy-univ.html