DALHOUSIE UNIVERSITY

Faculty of Science Course Syllabus
Department of Mathematics and Statistics
Introduction to data mining with R
STAT2450
Fall 2020 (08 Sept – 08Dec)

Instructor: Philippe Fullsack email: Philippe.Fullsack@dal.ca

Lectures: Lectures will associate pre-recorded videos and R markdown files, all posted on the BrightSpace course page. Additional documents (code notebooks, lecture slides, links to tutorial etc.) may be used to supplement the lectures.

This course includes no synchronous component.

Course Description

This course provides an introduction to data mining and R programming, suited for science students. Data mining methods include a vast set of tools developed in different areas for identifying the patterns in data. Students will learn programming methods for manipulating and exploring data through learning the basic ideas of some clustering, regression and classification methods. No prior programming knowledge is assumed.

Course Prerequisites

MATH 1000.03 or MATH 1215.03 and either (STAT 1060.03 or MATH 1060.03) or (STAT 2060.03 or MATH 2060.03) or DISP

Course Objectives/Learning Outcomes

The broad goals of this course are twofold:

Firstly, to teach students R programming and some general scientific computing methods. Roughly the first half of course will be allocated to R programming, including: using R as a calculator, data types, data structures, external files, loops and flow control, conditional execution, user defined functions, and use of built in statistical/graphical functions.

Secondly, to introduce a number of concepts for statistical learning, including: multiple regression, CART, supervised vs unsupervised learning, the bias variance trade-off, performance evaluation, cross validation, and bootstrapping.

Course Materials

There is a Brightspace site for the course. This is where assignment information and announcements will be posted. The Brightspace site may also contains links to documents and additional activities placed on the web server of the Department of Mathematics and Statistics.

Lecture videos will be posted as external learning resources on Brightspace. The Panopto software will be used to record/edit videos.

DALHOUSIE UNIVERSITY

Students will be required to use statistical software as part of this course. The software used in the course is the state-of-the-art open-source statistical package R. R is available from www.r-project.org for Mac OS, Windows, and Linux. An online environment for R is also available at rstudio.cloud.

- Required Textbook: "Introduction to Statistical Learning with Applications in R"
  by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, published by Springer.
  Hard copy in bookstore, or available free online at:

  https://www-bcf.usc.edu/~gareth/ISL/index.html
- Assignments will require use of the statistical software package R, together with the program RStudio, and the Rmarkdown library.
  Directions for installing the software are at:
  https://mathstat.dal.ca/~fullsack/stat2450/Notes/Rintro.html

Communication with Students

The course will make use of discussion forums on Brightspace to allow students to post their questions. Discussion lists will be organized by topics. Chats and forms will be used for additional interactions.

Course Organization

The course is divided in 7 modules. Each module includes 2 lectures, one quiz and one assignment. Modules will be delivered on a weekly to bi-weekly basis:

Component --	**TOPIC --**	**Start**	**End**
Module 1	R/Rstudio basics	Tue 8 Sep	Mon 21 Sep
Module 2	R: functions/distributions	Tue 22 Sep	Mon 28 Sep
Module 3	t-confidence interval	Tue 29 Sep	Mon 5 Oct
Module 4	Bootstrap	Tue 6 Oct	Mon 26 Oct
Module 5	Regression+Bootstrap	Tue 27 Oct	Mon 9 Nov
Module 6	Tree-based models	Tue 10 Nov	Mon 23 Nov
Module 7	Random forests	Tue 24 Nov	Mon 7 Dec

Course Assessment

The course will make use of two types of components for evaluating the progress of students: quizzes and assignments.

Quizzes and assignments will be available on Brightspace. All dates and times refer to those displayed in Brightspace. Note that dates will be set to Halifax local time.

Students located in another time zone will have to use the time displayed in Brightspace, not their local civil time.

DALHOUSIE UNIVERSITY

Assignments will be posted in R markdown format and students will be required to knit them to pdf in Rstudio. They will be marked by the TA of this course, based on solutions provided by me. I will eliminate the worst quiz on an individual basis.

Quizzes will be marked automatically on Brighspace.

Component --	**Weight (% of final grade) --**	**Available**	**Due Date**
Assignment 1	10 %	Tue 15 Sep	Tue 22 Sep
Quiz 1	5 %	Tue 15 Sep	Tue 22 Sep
Assignment 2	10 %	Tue 22 Sep	Tue 29 Sep
Quiz 2	5 %	Tue 22 Sep	Tue 29 Sep
Assignment 3	10 %	Tue 29 Sep	Tue 6 Oct
Quiz 3	5 %	Tue 29 Sep	Tue 6 Oct
Assessment 4	10 %	Tue 13 Oct	Tue 20 Oct
Quiz 4	5 %	Tue 13 Oct	Tue 20 Oct
Assessment 5	10 %	Tue 27 Oct	Tue 3 Nov
Quiz 5	5 %	Tue 27 Oct	Tue 3 Nov
Assessment 6	10 %	Tue 10 Nov	Tue 17 Nov
Quiz 6	5 %	Tue 10 Nov	Tue 17 Nov
Assessment 7	10 %	Tue 17 Nov	Tue 1 Dec
Quiz 7	5 %	Tue 17 Nov	Tue 1 Dec

Conversion of numerical grades to Final Letter Grades follows the Dalhousie Common Grade Scale:

A+ (90-100)	B+ (77-79)	C+ (65-69)	D (50-54)
A (85-89)	B (73-76)	C (60-64)	F (<50)
A- (80-84)	B- (70-72)	C- (55-59)	

Course Policies

Please note that all dates used in this course refer to the dates displayed in Brightspace.

This course follows the university policy on "missed or late academic requirements due to student absence" for assessments:

https://www.dal.ca/dept/university_secretariat/policies/academic/missed-or-late-academic-requirements-due-to-student-absence.html

Students experiencing a short-term absence which results in an assessment being missed must do the following:

- Contact the instructor(s) by e-mail prior to the scheduled due date of the assessment,
- Complete a Student Declaration of Absence Form within 3 calendar days of the last day of absence.

Late assignments will not be accepted.

It is expected that each student will write up their assignment independently.

Course Content

Listed below in roughly chronological order are the topics to be covered. Note that these may be altered slightly as the term progresses.

- Presentation of R and Rstudio
- R as a calculator, data types, data frames, importing and exporting data
- Controlling the program flow in R
- User-defined functions in R
- Built-in functions for classical probability distributions
- Exploratory data analysis in R, plots and graphs
- Simulation of coverage of t-confidence interval
- Introduction to linear and polynomial regression
- Introduction to the bootstrap method
- Bootstrap confidence intervals for the slope of a linear regression
- Elementary mathematical analysis of the bootstrap
- Introduction to tree-based regression: CART model
- Model validation techniques: leave-one-out and cross-validation
- Introduction to the Random Forests model
- Case studies: examples of applications of models to real data sets.

DALHOUSIE UNIVERSITY

Faculty of Science Course Syllabus Fall 2020 (revised June 2020)

Department of Mathematics and Statistics

Introduction to Data mining with R & STAT2450

University Policies and Statements

This course is governed by the academic rules and regulations set forth in the University Calendar and by Senate

Missed or Late Academic Requirements due to Student Absence

As per Senate decision instructors must not require medical notes of students who must miss an academic requirement, including a final exam, for courses offered during spring or summer sessions 2020 (until Aug 31, 2020).

Information on regular policy, including the use of the Student Declaration of Absence can be found here: **https://www.dal.ca/dept/university_secretariat/policies/academic/missed-or-late-academic-requirements-due-to-student-absence.html**

Academic Integrity

At Dalhousie University, we are guided in all of our work by the values of academic integrity: honesty, trust, fairness, responsibility and respect (The Center for Academic Integrity, Duke University, 1999). As a student, you are required to demonstrate these values in all of the work you do. The University provides policies and procedures that every member of the university community is required to follow to ensure academic integrity.

Information: https://www.dal.ca/dept/university_secretariat/academic-integrity.html

Accessibility

The Advising and Access Services Centre is Dalhousie's centre of expertise for student accessibility and accommodation. The advising team works with students who request accommodation as a result of a disability, religious obligation, or any barrier related to any other characteristic protected under Human Rights legislation (Canada and Nova Scotia).

Information: https://www.dal.ca/campus_life/academic-support/accessibility.html

Student Code of Conduct

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner—perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution.

Code: https://www.dal.ca/dept/university_secretariat/policies/student-life/code-of-student-conduct.html

Diversity and Inclusion – Culture of Respect

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality. Dalhousie is strengthened in our diversity. We are a respectful and inclusive

community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness

Statement: http://www.dal.ca/cultureofrespect.html

Recognition of Mi'kmaq Territory

Dalhousie University would like to acknowledge that the University is on Traditional Mi'kmaq Territory. The Elders in Residence program provides students with access to First Nations elders for guidance, counsel and support. Visit or e-mail the Indigenous Student Centre (1321 Edward St) (elders@dal.ca).

Information: https://www.dal.ca/campus_life/communities/indigenous.html

Important Dates in the Academic Year (including add/drop dates)

https://www.dal.ca/academics/important_dates.html

University Grading Practices

https://www.dal.ca/dept/university_secretariat/policies/academic/grading-practices-policy.html

Student Resources and Support

Advising

General Advising https://www.dal.ca/campus_life/academic-support/advising.html

Science Program Advisors: https://www.dal.ca/faculty/science/current-students/academic-advising.html

Indigenous Student Centre: https://www.dal.ca/campus_life/communities/indigenous.html

Black Students Advising Centre: https://www.dal.ca/campus_life/communities/black-student-advising.html

International Centre: https://www.dal.ca/campus_life/international-centre/current-students.html

Academic supports

Library: https://libraries.dal.ca/

Writing Centre: https://www.dal.ca/campus_life/academic-support/writing-and-study-skills.html

Studying for Success: https://www.dal.ca/campus_life/academic-support/study-skills-and-tutoring.html

Copyright Office: https://libraries.dal.ca/services/copyright-office.html

Fair Dealing Guidelines https://libraries.dal.ca/services/copyright-office/fair-dealing.html

Other supports and services

Student Health & Wellness Centre: https://www.dal.ca/campus_life/health-and-wellness/services-support/student-health-and-wellness.html

Student Advocacy: https://dsu.ca/dsas

DALHOUSIE UNIVERSITY

Ombudsperson: https://www.dal.ca/campus_life/safety-respect/student-rights-and-responsibilities/where-to-get-help/ombudsperson.html

Safety

Biosafety: https://www.dal.ca/dept/safety/programs-services/biosafety.html

Chemical Safety: https://www.dal.ca/dept/safety/programs-services/chemical-safety.html

Radiation Safety: https://www.dal.ca/dept/safety/programs-services/radiation-safety.html

Scent-Free Program: https://www.dal.ca/dept/safety/programs-services/occupational-safety/scent-free.html