CSC 422/522 Course Syllabus

CSC 422/522 – Automated Learning and Data Analysis

Section 001
Spring 2022
3 Credit Hours

Course Description

This course provides an introduction to concepts and methods for extracting knowledge or other useful forms of information from data. This activity, also known under names including data mining, knowledge discovery, and exploratory data analysis, plays an important role in modern science, engineering, medicine, business, and government. Students will apply supervised and unsupervised automated learning methods to extract patterns, make predictions and identify groups from data. Students will also learn about the overall process of data collection and analysis that provides the setting for knowledge discovery, and concomitant issues of privacy and security. Examples and projects introduce the students to application areas including electronic commerce, information security, biology, and medicine. Students cannot get credit for both CSC 422 and CSC 522.

Learning Outcomes

Upon completion, the student will be able to:

- Identify and contrast the major types of data and data representations with clear examples;
- List and explain the problems arising in preparing data for analysis, and the methods for addressing these problems;
- List and explain representative benefits and dangers of automated learning and data analysis;
- Identify ethical issues in data analysis applications, such as the impacts of data bias;
- Implement and apply various methods for supervised and unsupervised automated learning (e.g. Decision Trees, KNN, Naive Bayes, ANNs Regression, Clustering);
- Compare the strengths, weaknesses, and prerequisites of automated learning techniques;
- Explain and contrast methods for evaluating the performance of automated learning algorithms (e.g. holdout, k-fold crossvalidation, and leave-one-out crossvalidation);
- Design a detailed plan of analysis for a realistic data set;
- Apply automated data analysis tools to carry out a data analysis plan;
- Motivate, justify, and qualify conclusions obtained from an analysis.

Course Structure

The coursework consists of lectures, readings, homework assignments, tests, and a term project.

- **Lectures**: Each class will have a lecture. Some of the lectures will depart from the text, either in content or in order. Some material will be covered only in lecture; other material will be covered only in assigned readings. Tests will include material from lecture and readings. Students are responsible for all material presented or discussed in lecture.

- **Readings** will generally be taken from the text by from textbook, with possible supplements from the literature.
- **Homework** will require students to apply lecture topics, through short answer and data analysis questions. Students will be allowed to work in **teams** of 2-3 on Homeworks, but part or all of the homeworks will be submitted individually.
- **Exams** will consist of a midterm and a final exam. Statements of test objectives will be provided prior to the examinations to indicate their scope. The midterm will cover roughly the first half of the course content. The final exam will cover the entire course, but with an emphasis on the last half of the course content.
- **Project**: Each student must complete an extended data analysis project. The expectation is that the project will consist of performing and reporting on a data analysis task of your own interest. Projects will be completed in **teams of 3-4 students**. Each group must submit a written project report on their analysis. Project analyses will require the student to use data analysis software such as Python, R, or Matlab, but the course will not teach the use of software packages in detail. Full project guidelines will be provided in a separate document.

Course Policies

Collaboration Policy: For team homeworks and the final project, team members may collaborate as specified in the assignment, but teams must NOT talk with *other* teams about their work. Unwarranted similarity in teams' work constitutes a serious breach of academic integrity with appropriate penalties.

Disputing Grades Policy: If you wish to dispute a grade on an assignment or exam, you must submit a request **in writing, within 1 week of receiving the grade**. Requests should be submitted via Gradescope (where applicable) or as private Piazza posts, and will be addressed by TAs.

Instructors

Thomas Price (twprice) - *Instructor*
Email: twprice@ncsu.edu
Web Page: go.ncsu.edu/twprice
Phone: 9195151286
Office Location: See schedule
Office Hours: Schedule: go.ncsu.edu/csc522oh
Yang Shi (yshi26) - *Instructor*
Email: yshi26@ncsu.edu
Phone: N/A
Office Location: See schedule
Office Hours: Schedule: go.ncsu.edu/csc522oh
Chengyuan Liu (cliu32) - *Teaching Assistant*
Email: cliu32@ncsu.edu
Phone: N/A
Office Location: See schedule
Office Hours: Schedule: go.ncsu.edu/csc522oh
Parvez Rashid (mrashid4) - *Teaching Assistant*
Email: mrashid4@ncsu.edu
Phone: N/A

Office Location: See schedule
Office Hours: Schedule: go.ncsu.edu/csc522oh

Course Meetings

Lecture

Days: TH
Time: 10:15am - 11:45am
Campus: Centennial
Location: James B Hunt Library, room 4106
This meeting is required.

Course Materials

Textbooks

Introduction to Data Mining - *Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, and Vipin Kumar*
Edition: 2nd
ISBN: 0133128903
Cost: $35 (e-book)
This textbook is required.

Expenses

None.

Materials

None.

Requisites and Restrictions

Prerequisites

CSC 226 or LOG 201, ST 370, MA 305 or MA 405
Students will find introductions to artificial intelligence very helpful, but it is not required.

Co-requisites

None.

Restrictions

None.

Transportation

This course will not require students to provide their own transportation.

Safety & Risk Assumptions

None.

Grading

Grade Components

Component	Weight	Details
Homework	30	Four homeworks covering class topics, including theory and data analysis.
Midterm Exam	15	Midterm exam covering approximately the first half of the course.
Final Exam	25	Final exam covering the whole course.
Project	30	Project milestones and deliverables, including a

		proposal, midterm report, and a final written report.	

Letter Grades

This Course uses Standard NCSU Letter Grading Scale

97	≤	A+	≤	100
93	≤	A	<	97
90	≤	A-	<	93
87	≤	B+	<	90
83	≤	B	<	87
80	≤	B-	<	83
77	≤	C+	<	80
73	≤	C	<	77
70	≤	C-	<	73
67	≤	D	<	70
63	≤	D	<	67
60	≤	D-	<	63
0	≤	F	<	60

Requirements for Credit-Only (S/U) Grading

Performance in research, seminar and independent study types of courses (6xx and 8xx) is evaluated as either "S" (Satisfactory) or "U" (Unsatisfactory), and these grades are not used in computing the grade point average. For credit only courses (S/U) the requirements necessary to obtain the grade of "S" must be clearly outlined.

Undergraduate Credit-Only: To earn a grade of "S", the student must have a final grade of 70 or higher

as defined in the Grade Components.

Requirements for Auditors (AU)

Information about and requirements for auditing a course can be found at
http://policies.ncsu.edu/regulation/reg-02-20-04.

Policies on Incomplete Grades

If an extended deadline is not authorized by the Graduate School, an unfinished incomplete grade will automatically change to an F after either (a) the end of the next regular semester in which the student is enrolled (not including summer sessions), or (b) by the end of 12 months if the student is not enrolled, whichever is shorter. Incompletes that change to F will count as an attempted course on transcripts. The burden of fulfilling an incomplete grade is the responsibility of the student. The university policy on incomplete grades is located at http://policies.ncsu.edu/regulation/reg-02-50-03. Additional information relative to incomplete grades for graduate students can be found in the Graduate Administrative Handbook in Section 3.17.G at http://www.ncsu.edu/grad/handbook/index.php

Late Assignments

Each student will be allowed 2 total late days for homeworks without penalty for the entire semester, as well as 1 late day for project milestones. For example, you may be late by 1 day on two different homework submissions or late by 2 days on one submission. Each project team will also be granted 1 late day, for use only on the project. Weekends and holidays are also counted as late days. Late submissions are automatically considered as using late days. Once those days are used, you will be penalized according to how late the assignment is:

- at least 1 minute late but less than 24 hours late: the assignment will receive 75% credit
- at least 24 hours late but less than 48 hours late: the assignment will receive 50% credit
- at least 48 hours late: the assignment will not be accepted.

Late days may *only* be used for homeworks and project milestones (e.g. proposal, midterm report). No other assignments (final project report, exam, etc.) will be accepted late.

Attendance Policy

For complete attendance and excused absence policies, please see
http://policies.ncsu.edu/regulation/reg-02-20-03

Attendance Policy

It is expected that you will attend all classes and come prepared to engage in activities and discussion, with any relevant readings completed. All students are responsible for all material or

instructions introduced in class, which may include, but are not limited to, course material from textbooks, material from other sources (including from outside of the available slides), changes to the schedule, etc. The instructor and TAs cannot cover missed material with students who were absent, and much of the material covered in class is not available online.

Absences Policy

Academic Policy and Regulation REG02.20.3 - Attendance Regulations lists valid circumstances for an excused absence. Please also note:

1. For non-emergency situations, you must request advance permission to miss a class. You must request permission as soon as possible to miss class. If you must miss a class then, in all situations where it is practical, communicate with the instructor before the class to get permission (examples: you are not feeling well and wish to miss the class; you have a planned trip; you have a court appearance).
2. For emergency situations where you cannot communicate with the instructor in advance, you must contact the instructor as soon as you return to school (within 1 week of returning). Emergency absences must be certified in writing. Emergency illness that was not discussed with the instructor before missing the class must be certified by a physician in writing.

Makeup Work Policy

A make-up exam or final project presentation will only be allowed if you have an excused absence from class, as defined above. Other assignments cannot be made up unless explicitly mentioned by the instructor.

Additional Excuses Policy

None.

Academic Integrity

Academic Integrity

Students are required to comply with the university policy on academic integrity found in the Code of Student Conduct found at http://policies.ncsu.edu/policy/pol-11-35-01
Examples of academic integrity violations include (but are **not** limited to) when a student:

- Represents the work of others as his or her own;
- Obtains assistance in any academic work from another individual in a situation in which the student is expected to perform independently;
- Gives assistance to another individual in a situation in which that individual is expected to perform independently;
- Offers false data in support of laboratory or field work.

Violations will be reported to the Office of Student Conduct, which may impose penalties beyond

those recommended by the instructor.

Academic Honesty
See http://policies.ncsu.edu/policy/pol-11-35-01 for a detailed explanation of academic honesty.

Honor Pledge
Your signature on any test or assignment indicates "I have neither given nor received unauthorized aid on this test or assignment."

Digital Course Components
Students may be required to disclose personally identifiable information to other students in the course, via digital tools, such as email or web-postings, where relevant to the course. Examples include online discussions of class topics, and posting of student coursework. All students are expected to respect the privacy of each other by not sharing or using such information outside the course.
Digital Course Components: Electronically hosted course components include: Moodle, Piazza, Google Docs

Accommodations for Disabilities
Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with the Disability Resource Office at Holmes Hall, Suite 304, 2751 Cates Avenue, Campus Box 7509, 919-515-7653. For more information on NC State's policy on working with students with disabilities, please see the Academic Accommodations for Students with Disabilities Regulation (NCSU REG 02.20.01) (https://policies.ncsu.edu/regulation/reg-02-20-01/)

Non-Discrimination Policy
NC State provides equal opportunity and affirmative action efforts, and prohibits all forms of unlawful discrimination, harassment, and retaliation ("Prohibited Conduct") that are based upon a person's race, color, religion, sex (including pregnancy), national origin, age (40 or older), disability, gender identity, genetic information, sexual orientation, or veteran status (individually and collectively, "Protected Status"). Additional information as to each Protected Status is included in NCSU REG 04.25.02 (Discrimination, Harassment and Retaliation Complaint Procedure). NC State's policies and regulations covering discrimination, harassment, and retaliation may be accessed at http://policies.ncsu.edu/policy/pol-04-25-05 or https://oied.ncsu.edu/divweb/. Any person who feels that he or she has been the subject of prohibited discrimination, harassment, or retaliation should contact the Office for Institutional Equity and Diversity (OIED) at 919-515-3148.

Course Schedule
NOTE: The course schedule is subject to change.

Lecture TH 10:15am - 11:45am — All Weeks — 01/09/2022 - 04/20/2022
See course schedule at: go.ncsu.edu/csc522-schedule