

DS 520: Data Mining

School of Technology & Computing

3 Credits, Graduate Course
Winter 2021

Access to the Internet is required.

All written assignments must be in Microsoft-Word-compatible formats.

See the library's APA Style Guide tutorial for a list of resources that can help you use APA style.

Faculty Information

Professional experience information for instructors is found under *Faculty Information* in the online course menu.

Contact Information

Contact information for instructors is found under *Faculty Information* in the online course menu.

Course Description

Data mining involves the processing, analysis, and presentation of data to gain valuable information. Emphasis is placed on preparing high-quality, appropriate data that is relevant to the problem being assessed. Topics include clustering, classification, regression, pattern mining, prediction, association, and outlier detection, with attention being given to various forms of data, including time-series data and web data. Students will develop an understanding that many of these concepts depend on the notion of data proximity.

Course Resources

Required and recommended resources to complete coursework and assignments are found on the course [Reading List](#). The reading list can be found under *Course Information* in Blackboard as well as from the library homepage (“Find Your Textbook” button).

Note: Required resources that must be purchased by the student are tagged “Purchase from a vendor of your choosing.” Required resources with a direct link, “Available through CityU Library”, are available electronically at no cost to students.

Students in Canada may purchase course resources from the [Canada Bookstore](#), and students outside the U.S. and Canada should contact their advisor or textbook coordinator for additional information.

Course Outcomes

As a result of this course, students will know or be able to do the following:

- Understand the metrics appropriate for comparing various kinds of data.
- Understand the importance of feature selection for clustering, classification, and regression.
- Apply the Apriori algorithm to a range of applications for pattern mining.
- Analyze the data that may benefit from the use of regression and classification models.
- Evaluate the criteria that might lead to the selection of one method over another.
- Create a data mining deliverable using appropriate algorithms.

Grading Scale

The grades earned for the course will be calculated using City University of Seattle's decimal grading system, found in the current University Catalog (<https://www.cityu.edu/catalog/>).

Grading rubrics with details on how each assignment will be graded are located under *Assignments* and/or in *My Grades* in the online course menu. Students should review the rubric for each assignment prior to completing their work in order to understand how it will be assessed.

OVERVIEW OF REQUIRED ASSIGNMENTS	% OF FINAL GRADE	POINTS
The Muddiest Point (MP)	5%	50 = 5 points * 10 modules
Concept Test (CT)	5%	50 = 5 points * 10 modules
Discussion Board (DB)	10%	100 = 10 points * 10 modules
Hands-On Skill (HOS)	20%	200 = 20 points * 10 modules
Programming Exercise (PE)	30%	300 = 30 points * 10 modules
Knowledge Check (KC)	10%	100 = 10 points * 10 modules
Team Project (TP)	20%	Proposal: 30 points Progress: 70 points Final Report: 70 points Final PPT: 30 points Subtotal: 200 points
TOTAL	100%	1,000 points

Course Assignments and Grading

The instructor will provide grading rubrics that provide more detail as to how this assignment will be graded.

The Muddiest Point (MP)

Before class, students are required to submit the Muddiest Point (MP) activity. The purpose of this activity is to stimulate student engagement. The instructor uses the MP to assess how students understood the required readings. The instructor also uses the MP to customize the lecture scope to implement Just-in-Time Teaching (JiTT). The MP consists of writing a brief reflective essay (<= 50 words) identifying the most confusing part (i.e., the MP) of the content covered in the upcoming module. If a student understood all concepts, the student needs to explain the most exciting aspect. There is one multiple-choice question from the required reading to demonstrate that the student understood the required readings.

Criteria	% of Grade
Participation	40%
Writing	40%
Accuracy	20%
TOTAL	100%

Concept Test (CT)

The instructor poses a problem based on key concepts of a lecture. After reflecting on the problem, students submit their response and the instructor review them without providing a correct answer. Students discuss their thought process and solution with a peer. Students then commit to an answer and re-submits their responses. Instructor reviews responses and thought processes with the correct answer.

Criteria	% of Grade
Engagement	100%
TOTAL	100%

Discussion Board (DB)

A student posts an answer to a weekly discussion topic in Discussion Board. The student also posts a response to two other students' posts by the end of each module. Comments and questions should be clear and thoughtful, with correct grammar, spelling, and punctuation. The instructor will grade the quality of your discussion postings on both content and response.

Components	% of Grade
Participation	50%
Writing	50%
TOTAL	100%

Hands-on Skill (HOS)

The instructor will assign hands-on skill exercises to a pair of students in class or individually online. Students pair up and practice exercises to learn specific programming languages, application programming interfaces (APIs), or tools related to the programming assignments or virtual labs. Two quizzes measure hands-on skills acquired.

Criteria	% of Grade
Skill Exercise	70%
Engagement	20%
Correctness	10%
TOTAL	100%

Programming Exercise (PE)

The students must individually perform the programming exercise. Programs must be executable and robust. Non-executable programs will not receive any credits. Programs should deliver correct answers on all valid input and produce comprehensible error messages on invalid input. Programs also run correctly on all test data given within an assigned period. Students should write programs that are easy for other people to read.

Criteria	% of Grade
Program Execution	40%
User Requirement	40%
Program Documentation	20%
TOTAL	100%

Knowledge Check (KC)

Weekly quizzes measure knowledge concepts acquired. Focus on the underlying principles and concepts rather than memorization to solve the quizzes.

Criteria	% of Grade
Correctness	100%
TOTAL	100%

Team Project (TP)

Teams consist of three to four students. Each team will use an instructor-approved topic relevant to the course.

The paper is to be between 6 and 7 pages. The required template for class submissions comes from international organizations, the Education Special Interest Group and the Computing Education + Information systems Applied Research. ([EDSIG/CONISAR](#)). The instructor may recommend teams submit their paper to conferences. Submissions are optional and will not impact the course grade. Additional revisions may be required after the course.

Three report templates and one presentation template are provided. The file name consists of team project number, team number, and the list of your team members. For example, "TP01 T03 Sam John Mark."

- TP01 for the proposal - "TP01 T0X Author1 Author2 Author3.docx"
- TP02 for the progress report - "TP02 T0X Author1 Author2 Author3.docx"
- TP03 for the final report - "TP03 T0X Author1 Author2 Author3.docx"
- TP04 for the final presentation slide - "TP04 T0X Author1 Author2 Author3.pptx"

As in any scholarly writing, students should not merely copy information from another author. Students should use evidence to support the contentions they have drawn from their findings and critically analyze related literature. In essence, each paper needs to be an analytical paper, not a summary of readings.

In addition, a team slide deck presentation is required.

- The presentation consists of 15+4 slides: 15 slides for content and 4 slides for cover, agenda, key reference, and Q&A.
- The PPT template is provided. Your team can change design and color.
- A presentation video (15 minutes) is required.
- A demo video (a maximum of 1-2 minutes) may be included. The demo time is included in the 15 minutes presentation.

Four submissions are required according to the following schedule:

- Proposal (1 page; 30 points) - Starting (Module 1) & Ending (Module 3)
- Progress Report (3-4 pages; 70 points; graded after the proposal has been submitted) - Starting (Module 4) & Ending (Module 7)
- Final Report (6-7 pages; 70 points; graded after the progress has been submitted) - Starting (Module 8) & Ending (Module 10)
- Final PPT (15+4 slides, 30 points; graded after the final report has been submitted) - Starting (Module 8) & Ending (Module 10)

Students are expected to use the assigned readings, videos, and other materials throughout the quarter. Students will need to utilize additional sources that were not assigned by the professor. While stylized after an industry report, nonetheless, students are expected to employ APA formatting of citations, footnotes, and bibliography. Students must cite the sources of all ideas, facts, and information used that are not their own, even if they have put the information into their own words. Failure to do so is plagiarism, although the oversight is unintentional. To avoid plagiarism, check <https://library.cityu.edu/howto/apa-writing/avoid-plagiarism/>.

Project Description: Implementing Data Mining Algorithms

Students demonstrate competency implementing the data mining algorithms on a medium- to large-sized dataset they selected.

Team Project (TP) Report

The student will provide a report formatted based on a template provided by the instructor. Students are required to improve the writing iteratively and incrementally every week. The revision will always happen during a quarter. Students will add new required sections to the existing paper every week.

The final report is the culmination of applied research and activities conducted throughout the quarter. The final report/paper provides a detailed problem and its solution likely to be encountered by a company or organization described in a case study supplied by the student.

Criteria	% of Grade
Structure	20%
Content	30%
Writing	30%
Reference	10%
Collaboration	10%
TOTAL	100%

Team Project (TP) Presentation

The student will report on the research outcomes, development, or other project efforts to an academically appropriate committee in a public forum. The nature of the presentation content will determine the specific makeup of the audience. The student will choose the format of the presentation, in consultation with the advisor. The layout and design must be appropriate and

adequate to represent the outcomes of the effort. While students must make some form of a visual presentation, the presentation of the results may include publishing in a refereed publication, publication in a trade or popular magazine or journal, broadcast in an appropriate medium, or, in exceptional cases, limited dissemination within a closed community.

Each presenter will have 15 minutes for presentation and 5 minutes for questions and answers. Each presenter must keep the total presentation time limit strictly.

Criteria	% of Grade
Structure	20%
Visual Presentation	30%
Verbal Quality & Engagement	30%
Team Collaboration	20%
TOTAL	100%

Course Policies

Late Assignments

Please email me if you believe you will be late turning in an assignment. I will work with you to ensure you succeed in class, however, late assignments I will authorize late submission in a case-by-case basis.

Participation

I expect all students to participate in the discussion board and complete all assignments in a timely manner.

Professional Writing

Assignments require error-free writing that uses Standard English conventions and logical flow of organization to address topics clearly, completely, and concisely. CityU requires the use of APA style.

University Policies

Students are responsible for understanding and adhering to all of City University of Seattle's academic policies. The most current versions of these policies can be found in the [University Catalog](#) that is linked from the CityU Web site.

Non-Discrimination & Prohibition of Sexual Harassment

City University of Seattle adheres to all federal, state, and local civil rights laws prohibiting discrimination in employment and education. The University is committed to ensuring that the

education environment is bounded by standards of mutual respect and safety and is free from discriminatory practices.

In the U.S., the University is required by Title IX of the Education Amendments of 1972 to ensure that all of its education programs and activities do not discriminate on the basis of sex/gender. Sex include sex, sex stereotypes, gender identity, gender expression, sexual orientation, and pregnancy or parenting status. Sexual harassment, sexual assault, dating and domestic violence, and stalking are forms of sex discrimination, which are prohibited under Title IX and by City University of Seattle policy. City University of Seattle also prohibits retaliation against any person opposing discrimination or participating in any discrimination investigation or complaint process internal or external to the institution. Questions regarding Title IX, including its application and/or concerns about noncompliance, should be directed to the Title IX Coordinator. For a complete copy of the policy or for more information, visit <https://my.cityu.edu/titleix> or contact the Title IX Coordinator.

In Canada, in compliance with the British Columbia Human Rights Code, the Alberta Human Rights Act, WorksafeBC, and the Workers' Compensation Board of Alberta, the University believes that its environment should at all times be supportive and respectful of the dignity and self-esteem of individuals. Discrimination, harassment and bullying conduct, whether through person to person behaviour or via electronic communications such as email or social media is not acceptable and will not be tolerated. As an educational institution, it is our responsibility to cultivate an environment of excellence, equity, mutual respect and to recognize the value and potential of every individual. The University will take all necessary steps to meet or exceed the requirements of the law to prevent discrimination, harassment and bullying. The Respectful Workplace Policy for the prevention of discrimination, harassment and bullying policy and procedure can be found at <https://www.cityu.edu/discover-cityu/about-cityu/> under the Policies section or at <https://www.cityuniversity.ca/about/>.

Religious Accommodations

City University of Seattle has a policy for accommodation of student absences or significant hardship due to reasons of faith or conscience, or for organized religious activities. The University's policy, including more information about how to request an accommodation, is available in the University Catalog and on the my.cityu.edu student portal. Accommodations must be requested by the 20% mark of this course (e.g. day 14 of a ten-week course, day 7 of a 5-week course) using the Religious Accommodations Request Form found on the student dashboard in the my.cityu.edu student portal.

Academic Integrity

Academic integrity in students requires the pursuit of scholarly activity that is free from fraud, deception and unauthorized collaboration with other individuals. Students are responsible for understanding CityU's policy on academic integrity and adhering to its standards in meeting all course requirements. A complete copy of this policy can be found in the [University Catalog](#) under *Student Rights and Responsibilities* on the page titled *Academic Integrity Policy*.

Attendance

Students taking courses in any format at the University are expected to be diligent in their studies and to attend class regularly.

Regular class attendance is important in achieving learning outcomes in the course and may be a valid consideration in determining the final grade. For classes where a physical presence is required, a student has attended if they are present at any time during the class session. For online classes, a student has attended if they have posted or submitted an assignment. A complete copy of this policy can be in the [University Catalog](#) under *Student Rights and Responsibilities* on the page titled *Attendance*.

Final Assignment Due Date

Final assignments for each class at CityU must be due on or before the final date of the course as indicated in the university's course information system. Due dates that extend beyond the final date of the course may negatively impact tuition funding for students.

Support Services

Disability Services Accommodations Statement

Students with a documented disability who wish to request academic accommodations are encouraged to contact Disability Support Services to discuss accommodation requests and eligibility requirements. Please contact Disability Support Services at disability@cityu.edu or 206.239.4752 or visit the [Disability Support Services](#) page in the my.cityu.edu portal. Confidentiality will be observed in all inquiries. Once approved, information about academic accommodations will be shared with course instructors.

Library Services

CityU librarians are available to help students find the resources and information they need to succeed in this course. Contact a CityU librarian through the [Ask a Librarian](#) service, or access [library resources and services online](#), 24 hours a day, seven days a week.

Smarthinking Tutoring

CityU students have access to free online tutoring offered through Smarthinking, including writing support, from certified tutors 24 hours a day, seven days a week. Contact CityU's Student Support Center at help@cityu.edu to request a user name and password.