

CSCI-4144 — Data Mining and Warehousing Course Syllabus

Instructor Information

Instructor:Frank RudziczOffice:Goldberg 435E-mail:frank@dal.caOffice Hours:F 10h-11h

Class Meeting Time: MW 16h05-17h25 Room No: Studley CHEM 125.

Course Homepage: dal.brightspace.com/d21/home/311782 **Teaching Assistants:** CSCI4144-2024-instructors@cs.dal.ca

.

Important Dates

See Academic dates & deadlines and course website for any changes.

8 January First lecture

22 January Last day to add CSCI 4144

2 February Assignment 1 due

19–23 February Winter study break – no lectures or office hours

16 February Assignment 2 due
15 March Assignment 3 due
3 April Assignment 4 due
3 April Last lecture
TBD April Final exam

Course Description

This course reviews main concepts in data mining and data warehouses including objectives, architectures, algorithms, implementations, and applications. The topics covered include operational information process, decision-oriented information process, data warehousing, online analytical processing (OLAP), clustering, and classification. Selected system tools for data mining and data warehousing are introduced.

Learning Outcomes

- · Understand how to clean raw data
- Understand how to integrate data from different sources
- Understand how to effectively reduce the amount of raw data
- Understand how to how to transform and discretize data
- Describe the components of a data warehouse
- Design software to complete online analytical process (OLAP) operations
- Design software to perform clustering operations
- Compare different clustering algorithms
- Design software to perform classification or prediction operations
- Compare different classification and prediction algorithms

Course Rationale

This course provides a bridge between database design and statistical and machine learning aspects of data science.

Class Format and Course Communication

- Content will be delivered using a combination of lectures and posted assignments.
- Students must ask the instructor permission before recording class lectures.
- Course announcements will be posted to the course Brightspace.
- TA and Instructor office hours will be posted. Additional help is available at the CS Learning Centre.

Evaluation Criteria

- 1. Assignments (60%)
 - Four assignments, with relative weightings as follows:

Assignment with highest mark	20%	language: Python
Assignment with 2^{nd} highest mark	16%	language: Python
Assignment with 3^{rd} highest mark	14%	language: Python
Assignment with 4^{th} highest mark	10%	language: Python

- Deadlines: due at 19h
- Assignments must be submitted electronically.
- No collaboration is permitted on the assignments.
- All assignments may be checked with the Rubber Gasket plagiarism detection software.
- A 10% deduction is applied to late homework one minute after the due time. Thereafter, an additional 10% of the original maximum is deducted every 24 hours up to 72 hours late at which time the homework will receive a mark of zero. No exceptions will be made except in emergencies, at the instructor's discretion.
- 2. Final Exam (40%)
 - Scheduled by the university.
 - Will cover all material in the course.

Notes

- A minimum C grade is required in this course if it is core to your FCS degree, or if it will be used as a prerequisite
 for a subsequent CSCI course.
- As of 2019, students who receive a grade lower than C in the same required CS course twice, will be dismissed.
- The grade conversion scale in Section 17.1 of the Academic Regulations, Undergraduate Calendar will be used.
- A 24-hour 'silence policy' will be in effect we do not guarantee that the instructors or TAs will respond to your request within 24 hours of an assignment's due time.
- It is up to the discretion of the instructor to use remote proctoring in online testing. Students may be required to download proctoring software onto their devices. Students who cannot meet system requirements for remote proctoring should contact the instructor for an alternate assessment. (Typical system requirements are: (i) Mac OS or Windows, (ii) a web-cam, and (iii) an internet connection.)

Student Declaration of Absence

The Student Declaration of Absence policy shall apply.

Final Exam Requirements

- Photo ID is required.
- Closed book.
- No dictionaries, notes, calculators, cell phones, or other aids allowed.

Academic Standards

Failure to properly attribute sources in your work will be treated as an academic standards issue and points may be deducted for not following citation requirements. For example, forgetting to quote text taken from other sources, including large language models, failure to include in-text citations, or a failure to include required information in the citations or references. Please see the resources on proper citation provided by the Dalhousie Writing Center (https://dal.ca.libguides.com/c.php?g=257176&p=5001261).

Please note that if it appears that the error was made with intent to claim other people's work as your own such as a lack of both citations and references, an allegation of plagiarism will be submitted to the Faculty Academic Integrity Officer, which could result in consequences such as a course failure.

Required Texts and Resources

Optional Jiawei Han, Micheline Kamber, and Jian Pei (2011) Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers, ISBN 978-0123814791.

Available free at "Science Direct" section of Dal Library.

See website for lecture slides, additional reading, and other materials. Additional assistance is available from the Student Learning Centre (2nd floor, Goldberg CS Building).

Tentative Schedule of Topics

- 1. Introduction [Lecture 1]
- 2. Probability and statistics [Lectures 2-3]
- 3. Preprocessing [Lectures 4-5]
- 4. Warehousing and OLAP [Lectures 6-8]
- 5. Data Cubes [Lectures 9-11]

- 1. Patterns and associations [Lectures 12-14]
- 2. Classification [Lectures 15-17]
- 3. Clustering [Lectures 18-20]
- 4. Visualization and Decisions [Lectures 21-23]

Prerequisites

CSCI-2110.03 and CSCI-2141.03.

Responsible Computing Policy

Usage of all computing resources in the Faculty of Computer Science must be within the Dalhousie Acceptable Use Policies (https://www.dal.ca/dept/university_secretariat/policies/information-management-and-tacceptable-use-policy-.html) and the Faculty of Computer Science Responsible Computing Policy. (https://www.dal.ca/content/dam/dalhousie/pdf/faculty/computerscience/policies-procedures/fcs_policy_local.pdf)

Use of Plagiarism Detection Software

All submitted assignment may be passed through a plagiarism detection software, such as the Moss Software Similarity Detection System (https://theory.stanford.edu/~aiken/moss/), or similar systems. If a student does not wish to have their assignments passed through plagiarism detection software, they should contact the instructor for an alternative. Please note, that code not passed through plagiarism detection software will necessarily receive closer scrutiny. https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf

Use of Artificial Intelligence Tools

You may use AI-driven tools to assist you in learning but remember that your objective is to understand, achieve, and apply the course competencies and outcomes. While you may use tools for learning, specific assessments in this course will disallow the use of AI-driven tools to assert that you have attained course learning outcomes. This is because a graduate must be able to analyze, assess and produce work unassisted by AI technology. Where tools are allowed: you must acknowledge all tools used to assist you. If applicable, you must provide links to chat logs. Using AI-driven tools where prohibited constitutes an academic offence.

Culture of Respect

Every person has a right to respect and safety. We believe inclusiveness is fundamental to education and learning. Misogyny and other disrespectful behaviour in our classrooms, on our campus, on social media, and in our community is unacceptable. As a community, we must stand for equality and hold ourselves to a higher standard.

What we all need to do 1:

- 1. **Be Ready to Act:** This starts with promising yourself to speak up to help prevent it from happening again. Whatever it takes, summon your courage to address the issue. Try to approach the issue with open-ended questions like "Why did you say that?" or "How did you develop that belief?"
- 2. **Identify the Behaviour:** Use reflective listening and avoid labeling, name-calling, or assigning blame to the person. Focus the conversation on the behaviour, not on the person. For example, "The comment you just made sounded racist, is that what you intended?" is a better approach than "You're a racist if you make comments like that"
- 3. Appeal to Principles: This can work well if the person is known to you, like a friend, sibling, or co-worker. For example, "I have always thought of you as a fair-minded person, so it shocks me when I hear you say something like that."
- 4. **Set Limits:** You cannot control another person's actions, but you can control what happens in your space. Do not be afraid to ask someone "Please do not tell racist jokes in my presence anymore" or state "This classroom is not a place where I allow homophobia to occur." After you have set that expectation, make sure you consistently maintain it.
- 5. **Find or be an Ally:** Seek out like-minded people that support your views, and help support others in their challenges. Leading by example can be a powerful way to inspire others to do the same.
- Be Vigilant: Change can happen slowly, but do not let this deter you. Stay prepared, keep speaking up, and do not let yourself be silenced.

¹Source: Speak Up! ©2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full "Speak Up" document found at: http://www.dal.ca/dept/dalrespect.html Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University 902.494.4140 lyndsay.anderson@dal.ca www.dal.ca/think.

University Statements

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

Territorial Acknowledgement

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

Internationalization

At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders."

https://www.dal.ca/about-dal/internationalization.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. 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.

(read more: http://www.dal.ca/dept/university_secretariat/academic-integrity.html)

Accessibility

The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion please contact:

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

for all courses offered by Dalhousie with the exceptionof Truro.

Conduct in the Classroom — Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

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 (Strategic Priority 5.2).

(read more:http://www.dal.ca/cultureofrespect.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.

(read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-reposite Code%20of%20Student%20Conduct%20rev%20Sept%202021.pdf)

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. (read more:

https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.
html)

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. (read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf)

Student Use of Course Materials

These course materials are designed for use as part of the CSCI courses at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading material to a commercial third party website) may lead to a violation of Copyright law.

Learning and Support Resources

Please see https://www.dal.ca/campus_life/academic-support.html



CSCI-6405 — Data Mining and Warehousing Course Syllabus

Instructor Information

Instructor:Frank RudziczOffice:Goldberg 435E-mail:frank@dal.caOffice Hours:F 10h-11h

Class Meeting Time: MW 16h05-17h25 Room No: Studley CHEM 125.

Course Homepage: dal.brightspace.com/d21/home/311782 **Teaching Assistants:** CSCI4144-2024-instructors@cs.dal.ca

.

Important Dates

See Academic dates & deadlines and course website for any changes.

8 January First lecture

22 January Last day to add CSCI 6405

19–23 February Winter study break – no lectures or office hours

3 April Final project report due

3 April Last lecture TBD April Final exam

Course Description

This course reviews main concepts in data mining and data warehouses including objectives, architectures, algorithms, implementations, and applications. The topics covered include operational information process, decision-oriented information process, data warehousing, online analytical processing (OLAP), clustering, and classification. Selected system tools for data mining and data warehousing are introduced.

Learning Outcomes

- · Understand how to clean raw data
- · Understand how to integrate data from different sources
- Understand how to effectively reduce the amount of raw data
- Understand how to how to transform and discretize data
- Describe the components of a data warehouse
- Design software to complete online analytical process (OLAP) operations
- Design software to perform clustering operations
- Compare different clustering algorithms
- Design software to perform classification or prediction operations
- Compare different classification and prediction algorithms

Course Rationale

This course provides a bridge between database design and statistical and machine learning aspects of data science.

Class Format and Course Communication

- Content will be delivered using a combination of lectures and posted assignments.
- Students must ask the instructor permission before recording class lectures.
- Course announcements will be posted to the course Brightspace.
- TA and Instructor office hours will be posted. Additional help is available at the CS Learning Centre.

Evaluation Criteria

- 1. Projects (60%)
 - · One project, as follows:

Project 60% language: e.g., Python

- Projects must be submitted electronically.
- Projects can be completed in teams of 1 or 2 graduate students.
- All projects may be checked with the Rubber Gasket plagiarism detection software.
- 2. Final Exam (40%)
 - · Scheduled by the university.
 - · Will cover all material in the course.

Notes

- A minimum B- grade is required to pass this course
- The grade conversion scale in Section 7.7.1 of the Faculty of Graduate Studies Regulations in the Graduate calendar will be used. https://academiccalendar.dal.ca/~/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=124&topicgroupid=35472
- A 24-hour 'silence policy' will be in effect we do not guarantee that the instructors or TAs will respond to your request within 24 hours of an assignment's due time.
- It is up to the discretion of the instructor to use remote proctoring in online testing. Students may be required to download proctoring software onto their devices. Students who cannot meet system requirements for remote proctoring should contact the instructor for an alternate assessment. (Typical system requirements are: (i) Mac OS or Windows, (ii) a web-cam, and (iii) an internet connection.)

Student Declaration of Absence

The Student Declaration of Absence policy shall apply.

Final Exam Requirements

- Photo ID is required.
- · Closed book.
- No dictionaries, notes, calculators, cell phones, or other aids allowed.

Academic Standards

Failure to properly attribute sources in your work will be treated as an academic standards issue and points may be deducted for not following citation requirements. For example, forgetting to quote text taken from other sources, including large language models, failure to include in-text citations, or a failure to include required information in the citations or references. Please see the resources on proper citation provided by the Dalhousie Writing Center (https://dal.ca.libguides.com/c.php?g=257176&p=5001261).

Please note that if it appears that the error was made with intent to claim other people's work as your own such as a lack of both citations and references, an allegation of plagiarism will be submitted to the Faculty Academic Integrity Officer, which could result in consequences such as a course failure.

Required Texts and Resources

Optional Jiawei Han, Micheline Kamber, and Jian Pei (2011) Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers, ISBN 978-0123814791.

Available free at "Science Direct" section of Dal Library.

See website for lecture slides, additional reading, and other materials. Additional assistance is available from the Student Learning Centre (2nd floor, Goldberg CS Building).

Tentative Schedule of Topics

- 1. Introduction [Lecture 1]
- 2. Probability and statistics [Lectures 2-3]
- 3. Preprocessing [Lectures 4-5]
- 4. Warehousing and OLAP [Lectures 6-8]
- 5. Data Cubes [Lectures 9-11]

- 1. Patterns and associations [Lectures 12-14]
- 2. Classification [Lectures 15-17]
- 3. Clustering [Lectures 18-20]
- 4. Visualization and Decisions [Lectures 21-23]

Responsible Computing Policy

Usage of all computing resources in the Faculty of Computer Science must be within the Dalhousie Acceptable Use Policies (https://www.dal.ca/dept/university_secretariat/policies/information-management-and-tacceptable-use-policy-.html) and the Faculty of Computer Science Responsible Computing Policy. (https://www.dal.ca/content/dam/dalhousie/pdf/faculty/computerscience/policies-procedures/fcs_policy_local.pdf)

Use of Plagiarism Detection Software

All submitted assignment may be passed through a plagiarism detection software, such as the Moss Software Similarity Detection System (https://theory.stanford.edu/~aiken/moss/), or similar systems. If a student does not wish to have their assignments passed through plagiarism detection software, they should contact the instructor for an alternative. Please note, that code not passed through plagiarism detection software will necessarily receive closer scrutiny. https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf

Use of Artificial Intelligence Tools

You may use AI-driven tools to assist you in learning but remember that your objective is to understand, achieve, and apply the course competencies and outcomes. While you may use tools for learning, specific assessments in this course will disallow the use of AI-driven tools to assert that you have attained course learning outcomes. This is because a graduate must be able to analyze, assess and produce work unassisted by AI technology. Where tools are allowed: you must acknowledge all tools used to assist you. If applicable, you must provide links to chat logs. Using AI-driven tools where prohibited constitutes an academic offence.

Culture of Respect

Every person has a right to respect and safety. We believe inclusiveness is fundamental to education and learning. Misogyny and other disrespectful behaviour in our classrooms, on our campus, on social media, and in our community is unacceptable. As a community, we must stand for equality and hold ourselves to a higher standard.

What we all need to do 1:

- 1. **Be Ready to Act:** This starts with promising yourself to speak up to help prevent it from happening again. Whatever it takes, summon your courage to address the issue. Try to approach the issue with open-ended questions like "Why did you say that?" or "How did you develop that belief?"
- 2. **Identify the Behaviour:** Use reflective listening and avoid labeling, name-calling, or assigning blame to the person. Focus the conversation on the behaviour, not on the person. For example, "The comment you just made sounded racist, is that what you intended?" is a better approach than "You're a racist if you make comments like that."
- 3. **Appeal to Principles:** This can work well if the person is known to you, like a friend, sibling, or co-worker. For example, "I have always thought of you as a fair-minded person, so it shocks me when I hear you say something like that."
- 4. Set Limits: You cannot control another person's actions, but you can control what happens in your space. Do not be afraid to ask someone "Please do not tell racist jokes in my presence anymore" or state "This classroom is not a place where I allow homophobia to occur." After you have set that expectation, make sure you consistently maintain it.
- 5. **Find or be an Ally:** Seek out like-minded people that support your views, and help support others in their challenges. Leading by example can be a powerful way to inspire others to do the same.
- 6. **Be Vigilant:** Change can happen slowly, but do not let this deter you. Stay prepared, keep speaking up, and do not let yourself be silenced.

University Statements

This course is governed by the academic rules and regulations set forth in the University Calendar and the Senate. https://academiccalendar.dal.ca/Catalog/ViewCatalog.aspx?pageid=viewcatalog&catalogid=117&loaduseredits=False

Territorial Acknowledgement

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

Internationalization

At Dalhousie, 'thinking and acting globally' enhances the quality and impact of education, supporting learning that is "interdisciplinary, cross-cultural, global in reach, and orientated toward solving problems that extend across national borders."

https://www.dal.ca/about-dal/internationalization.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. 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.

(read more: http://www.dal.ca/dept/university_secretariat/academic-integrity.html)

¹Source: Speak Up! ©2005 Southern Poverty Law Center. First Printing. This publication was produced by Teaching Tolerance, a project of the Southern Poverty Law Center. Full "Speak Up" document found at: http://www.dal.ca/dept/dalrespect.html Revised by Susan Holmes from a document provided April 2015 by Lyndsay Anderson, Manager, Student Dispute Resolution, Dalhousie University 902.494.4140 lyndsay.anderson@dal.ca www.dal.ca/think.

Accessibility

The Student Accessibility Centre is Dalhousie's centre of expertise for matters related to student accessibility and accommodation. If there are aspects of the design, instruction, and/or experiences within this course (online or in-person) that result in barriers to your inclusion please contact:

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

for all courses offered by Dalhousie with the exceptionof Truro.

Conduct in the Classroom — Culture of Respect

Substantial and constructive dialogue on challenging issues is an important part of academic inquiry and exchange. It requires willingness to listen and tolerance of opposing points of view. Consideration of individual differences and alternative viewpoints is required of all class members, towards each other, towards instructors, and towards guest speakers. While expressions of differing perspectives are welcome and encouraged, the words and language used should remain within acceptable bounds of civility and respect.

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 (Strategic Priority 5.2).

(read more:http://www.dal.ca/cultureofrespect.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.

(read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-reposito Code%20of%20Student%20Conduct%20rev%20Sept%202021.pdf)

Fair Dealing Policy

The Dalhousie University Fair Dealing Policy provides guidance for the limited use of copyright protected material without the risk of infringement and without having to seek the permission of copyright owners. It is intended to provide a balance between the rights of creators and the rights of users at Dalhousie. (read more: https://www.dal.ca/dept/university_secretariat/policies/academic/fair-dealing-policy-.html)

Originality Checking Software

The course instructor may use Dalhousie's approved originality checking software and Google to check the originality of any work submitted for credit, in accordance with the Student Submission of Assignments and Use of Originality Checking Software Policy. Students are free, without penalty of grade, to choose an alternative method of attesting to the authenticity of their work, and must inform the instructor no later than the last day to add/drop classes of their intent to choose an alternate method. (read more: https://cdn.dal.ca/content/dam/dalhousie/pdf/dept/university_secretariat/policy-repository/OriginalitySoftwarePolicy.pdf)

Student Use of Course Materials

These course materials are designed for use as part of the CSCI courses at Dalhousie University and are the property of the instructor unless otherwise stated. Third party copyrighted materials (such as books, journal articles, music, videos, etc.) have either been licensed for use in this course or fall under an exception or limitation in Canadian Copyright law. Copying this course material for distribution (e.g. uploading material to a commercial third party website) may lead to a violation of Copyright law.

Learning and Support Resources

Please see https://www.dal.ca/campus_life/academic-support.html