DSCI 4520 – Introduction to Data Mining
Fall 2023

“What I cannot create, I do not understand.”
Richard Feynman

Course Information
Course: DSCI 4520.003
Course Name: Introduction to Data Mining
Class Location: BLB 035
Time: Mo 6:30 pm – 9:20 pm
Website: http://learn.unt.edu

Instructor Contact Information
Instructor: Dr. Mahdi Ahmadi
Office: BLB 331C
Phone: 940-565-2946
Email: Mahdi.Ahmadi@unt.edu
Office Hours: Mo 10:00 am – 12:00 pm
Th 10:00 am – 12:00 pm
or by appointment
Zoom link: https://unt.zoom.us/my/mahdi.ahmadi

Course Description
How can data be utilized to make better decisions and solve real-life problems? How can data-driven information shape and support business decisions? What is the true value of data for businesses? How to turn data into solutions for business problems? This course is designed to address these and similar questions by focusing on extracting relevant and useful insights from small or large amounts of data and creating predictive models through data mining techniques. In this course, we introduce technologies, processes, and practices to explore and investigate data to gain insights for better business outcome. The course will start with concepts and techniques for data preparation, summarization, visualization, and statistical analysis. Then, we will introduce supervised and unsupervised methods such as clustering, linear regression, logistic regression, decision trees, neural networks, association rule extraction, and classification ensembles to discover and predict relationships and patterns in data. This course will examine principles and techniques of predictive analytics that can be used to shape the present and future of businesses.
Learning Objectives
1- To learn basic concepts and techniques of data mining such as exploratory analysis, data pre-
processing, linear regression models, classification models, clustering, model performance
evaluation, and neural networks.
2- To understand business problems, transform them into data mining problems, solve the
problems, interpret results in the business context, and present the solutions effectively.
3- To identify when and how each data mining technique can be used to solve business problems.
4- To acquire a practical understanding of preparing and evaluating predictive models and
explorative algorithms

Textbook, Materials, and Resources

Textbook
- Predictive Analytics and Data Mining Concepts and Practice with RapidMiner. Vijay Kotu, and

Note: You DO NOT have to buy this book. An electronic version of this book is available for UNT students
through the UNT Library resources: https://discover.library.unt.edu/catalog/b6041002

I may also post educational videos, scholarly papers, blog posts, articles, and other materials on Canvas.
Students should check their Canvas account regularly for additional materials.

Software
This is a hands-on course and students will learn data mining techniques by practicing them. We will be
using the following software:

- Excel for statistical analysis, data exploration, and practicing data mining concepts
- Tableau for data visualization
- RapidMiner for building and evaluating data mining models.

Tableau and RapidMiner are available on the Ryan College of Business virtual windows machines through
UNT MyLab.

RapidMiner offers a free educational license, and therefore, students can install it on their own
computers. I recommend this because the virtual computers might not be available all the time.
Detailed instructions on how to access, install, and activate RapidMiner on personal computers will be
available on Canvas.

Generative AI (ChatGPT, Google Bard, etc.)
Generative AI is one of the most important technological innovations of our time and will change the
future of data analytics, business processes, education, and economy. In this course you will have
opportunities to practice working with ChatGPT (or similar platforms) to solve data mining problems. I
will give you specific instructions on where and for what purpose to use ChatGPT for the assignments.
Delivery Method and Technology Requirements
The course will be delivered in person. Active class participation is crucial to the learning and success of the students. Active participation means participation in the class practices and discussions, presenting your team project in class, and engaging in constructive dialogues. I will deliver a lecture in each class, and the rest of the class will be discussions, students’ presentations, case studies, and hands-on practices. We will rely heavily on Canvas for the delivery of materials, practices, assignments, exams, and projects. It is students' responsibility to check their Canvas account and email regularly.

A laptop is required during all classes and exams. We will use a variety of analytical and data tools in the class. It is the student’s responsibility to have a working and functional laptop during class time and exams. Laptops can be checked out from the UNT Library.

During my office hours I will be available both in my office and on Zoom. Students can ask for an appointment outside of office hours. All communications with me should ONLY be done through my official UNT email address or Canvas.

Attendance Policy
Class attendance is required for all students for all class meetings. Students are expected to actively participate in all classes. Students are strongly encouraged to ask relevant questions during my lectures and other students’ presentations. Students are required to abide by the attendance policy established for the course. It is important that you communicate with the professor and the instructional team prior to being absent, so you, the professor, and the instructional team can discuss and mitigate the impact of the absence on your attainment of course learning goals. Please inform the professor and instructional team if you are unable to attend class meetings because you are ill, in mindfulness of the health and safety of everyone in our community.

Course Schedule
The following table shows the tentative schedule for the course. Dates, topics, and activities are subject to change.

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Subject</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21-Aug</td>
<td>Course syllabus + Introduction to Data Mining</td>
<td>Syllabus + Ch 1</td>
</tr>
<tr>
<td>2</td>
<td>28-Aug</td>
<td>Review of Business Statistics + Data Mining Process</td>
<td>Ch 2</td>
</tr>
<tr>
<td>3</td>
<td>4-Sep</td>
<td>Labor Day (university closed) -- Read the notes</td>
<td>Ch 3</td>
</tr>
<tr>
<td>4</td>
<td>11-Sep</td>
<td>Data Preparation, Visualization, and Exploratory Analytics II</td>
<td>Ch 3 &amp; 11</td>
</tr>
<tr>
<td>5</td>
<td>18-Sep</td>
<td>Clustering and Segmentation</td>
<td>Ch 7</td>
</tr>
<tr>
<td>6</td>
<td>25-Sep</td>
<td>k-Nearest Neighbors + Linear Regression I</td>
<td>Ch 5</td>
</tr>
<tr>
<td>7</td>
<td>2-Oct</td>
<td>Linear Regression II</td>
<td>Ch 5</td>
</tr>
<tr>
<td>8</td>
<td>9-Oct</td>
<td>Exam 1 - Project First Milestone</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>16-Oct</td>
<td>Introduction to Classification &amp; Evaluating Classifiers</td>
<td>Ch 8 &amp; 4</td>
</tr>
<tr>
<td>10</td>
<td>23-Oct</td>
<td>Naïve Bayes Classifier &amp; Logistic Regression</td>
<td>Ch 4</td>
</tr>
</tbody>
</table>
11 30-Oct  Decision Trees & Random Forests  Ch 4
12 6-Nov   Exam 2 - Project Second Milestone
13 13-Nov  Association Rules & Recommender Systems  Ch 6
14 20-Nov  Thanksgiving Break (no classes)
15 27-Nov  Neural Networks  Ch 4
16 4-Dec   Course Review and Project Presentations
17 11-Dec  Final exam

**IMPORTANT NOTE:**

There will be no face-to-face class on September 4th because the university is closed for Labor Day. However, I will record a lecture and instruction and post it on Canvas. The lecture and materials in this virtual session should be considered like all other educational materials in face-to-face meetings. I will teach the first part of the Data Preparation, Visualization, and Exploratory Analysis (chapter 3 of the textbook) in this virtual class. Students should watch the recorded video, do the exercises, and follow the instructions in the video and on Canvas.

---

**Grade Components**

Grades will be based on student performance in the course, exams, assignments, term project, and class participation.

<table>
<thead>
<tr>
<th>Assessment Component</th>
<th>Points</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice Assignments</td>
<td>300</td>
<td>30%</td>
</tr>
<tr>
<td>Quizzes</td>
<td>100</td>
<td>10%</td>
</tr>
<tr>
<td>Exams</td>
<td>350</td>
<td>35%</td>
</tr>
<tr>
<td>Term Project</td>
<td>250</td>
<td>25%</td>
</tr>
</tbody>
</table>

**Practice Assignments**

This is a hands-on course in which you would learn how to create, use, and interpret data mining techniques. Throughout the course, 5 practice assignments with multiple parts are designed to help you learn data mining and data analytics techniques and concepts. I will explain and initiate each assignment in the class and help students start working on them.

**Quizzes**

The purpose of quizzes is to assess and establish your learning in a short period and prepare you for the mid-term and final exams. All quizzes will be open-book and should be taken individually. There will be 5 quizzes throughout the semester.
Exams
Three closed-book and closed-note exams will be given (two exams in the middle of the semester and a final exam). The final exam constitutes 15% of your final grade and will be a comprehensive evaluation of everything you’ve learned in this course. The tentative dates of the exam are shown in the class schedule table. All exams should be taken in the classroom during the normal class meeting time through LockDown Browser. The final exam will be given on Dec 11 at 6:30 pm (according to the university Final Exam Schedule). Students will be individually tested over all materials covered in the lectures, course readings, assignments, case studies, and tutorials.

Make-up exams will NOT be given in general. If you have a UNT-authorized valid and verified excuse for missing an exam, please communicate it with me at least 48 hours before the exam date. Requests for a make-up AFTER the exam date will only be accepted to review within 48 hours and it requires extraordinary circumstances with valid and verified documents. No request will be reviewed after 48 hours.

Term Project
This is an opportunity for students to work in a team to solve a business problem with the skills they learn in this course. I will give a list of problems and datasets for the students to choose their topic and data. Students are welcome to bring in their own problem definition and dataset, but they must get my approval prior to starting work on it as their term project. There will be two milestones (indicated in the class schedule table) to assess and review the progress of the project. Detailed instructions and rubrics will be posted on Canvas after the semester starts.

Final Grades Policy

<table>
<thead>
<tr>
<th>Final point score out of 1000</th>
<th>Final letter grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤900</td>
<td>A</td>
</tr>
<tr>
<td>800-899</td>
<td>B</td>
</tr>
<tr>
<td>700-799</td>
<td>C</td>
</tr>
<tr>
<td>600-699</td>
<td>D</td>
</tr>
<tr>
<td>&lt; 600</td>
<td>F</td>
</tr>
</tbody>
</table>

Please read Instructor Policies and UNT Policies sections on academic integrity carefully.

Grade Disputes
Students have the right and are welcome to review assignments and exams after grading to improve their understanding of course material and check for the presence of grading errors. This review must be conducted within 7 days after the grades of the assignment or exam are posted on Canvas. All requests should be sent directly to my UNT email only with a clear and concise explanation of why and where you believe there has been a grading error. Grade appeals are to ensure mistakes do not negatively impact your grades. They are not intended to ensure you receive your desired final grade. Requests for blanket reconsideration of your graded deliverables will not be reviewed or answered. No changes to grades will be made after this period.
Instructor Policies

Class Code of Conduct
Student behaviors that interfere with the instructor’s ability to conduct a class or other students’ opportunity to learn are unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The university’s expectations for student conduct apply to all instructional forums, including university and electronic classrooms, labs, discussion groups, field trips, etc.

As a courtesy to the class, your classmates, and your instructor, you are asked to set your cell phones to vibrate. In the case of a personal emergency that requires you to answer your phone, you are asked to step out of the classroom.

Academic Integrity
Ryan G. College of Business and I take academic integrity and honesty extremely seriously for all course activities such as quizzes, assignments, exams, and the term project. This course is an excellent opportunity to practice what is expected from you as a business professional regarding integrity, trust, and honesty. All students are required to maintain the highest standards of ethical and professional conduct when taking exams and doing assignments and the term project.

In this course, unless otherwise stated, individual work should be completed alone and using only resources explicitly outlined in the instructions. While external research may be permitted in assignment instructions, utilizing resources such as Chegg, Course Hero, and similar websites is not appropriate and is expressly forbidden. What may appear to be a relatively minor step outside the bounds of acceptable behavior can have a monumental impact on success within your academic program and beyond.

Students caught cheating or plagiarizing will receive a zero for that particular assignment or exam for the first time. Engaging in such behavior for the second time will result in an F for the course, with absolutely no exception. Additionally, the incident may be reported to the Dean of Students, who may impose further penalties or sanctions ranging from admonition to expulsion from the university.

Using generative AI platforms for doing the assignments is allowed only if it is clearly cited and explained in the assignment by the professor. This means, if you are asked to use ChatGPT to do an assignment, you are allowed to do it and you must include all the prompts and answers in the submitted documents.

Read the definition of cheating by the UNT carefully: https://policy.unt.edu/policy/06-003

Communication
Canvas is the primary channel for announcements, changes in due dates or the syllabus, new materials, grades, etc. It is the student’s responsibility to check their Canvas account for updates.

If a student wants to initiate communication outside of Canvas, they should do it with their UNT email to my UNT email: Mahdi.Ahmadi@unt.edu. Emails from non-UNT email systems will not be read. You can expect my response within 24 hours. Virtual appointments will be held on Zoom. In your email subject like you must always mention course code + section (DSCI 4520.003) and in the body of the email you
should clearly and concisely explain what it is that you are asking. This is an opportunity for you to practice writing professional emails: be clear, polite, concise, and to the point.

I value your feedback greatly and I strongly encourage you to not hesitate to communicate with me about any issues that you may have or see in the course. To make it easier, I will create an **anonymous electronic drop box** where you can leave your suggestions, opinions, and questions anonymously. Details will be posted on Canvas.

**Changes in the Syllabus**
The materials presented in this syllabus are tentative. I reserve the right to change and improve the materials and requirements as the semester unfolds, with sufficient announcements concerning exams, and assignments.

**Late Work**
Only assignments that are submitted before the due date are considered to be graded without a deduction for late submission. If you submit an assignment within 24 hours of the due date, you will lose 20% of the full credit, and if you submit it after the first day and before the third day after the due date your grade starts from 50% of the full grade. After three days, you will receive a zero, with no exceptions, unless you have a university-authorized excused absence and provides verified documentation **24 hours before the deadline**.

---

**University Policies and Other Important Information**

**Academic Integrity Policy**
The University of North Texas promotes the integrity of learning and embraces the core values of trust and honesty. Academic integrity is based on educational principles and procedures that protect the rights of all participants in the educational process and validate the legitimacy of degrees awarded by the University. In the investigation and resolution of allegations of student academic dishonesty, the University’s actions are intended to be corrective, educationally sound, fundamentally fair, and based on reliable evidence. Please read the full text of the UNT policy [here](#).

**Disability Accommodation Policy**
The University of North Texas (UNT or University) does not discriminate on the basis of disability in admission, treatment, or access to its programs or activities, nor in employment in its programs or activities. The University is committed to providing equal educational access for qualified students with disabilities in accordance with state and federal laws, including the Americans with Disabilities Act of 1990 as Amended, and Section 504 of the Rehabilitation Act of 1973. In addition, the University is committed to making all programs and activities sponsored by UNT accessible, as required by the Texas Accessibility Standards and the Americans with Disabilities Act Accessibility Guidelines. To this end, all academic units are willing to make reasonable and appropriate adjustments to the classroom environment and the teaching, testing, or learning methodologies in order to facilitate equality of educational access for persons with disabilities.

Please read the full text of the UNT policy [here](#). You can learn more about the UNT Office of Disability Access on their [website](#) or through their email ([APPLY.ODA@UNT.EDU](mailto:APPLY.ODA@UNT.EDU)) or their phone number: 940-565-4323
Academic Deadlines

Dates of drop deadlines, final exams, etc., are published in the university calendar and the schedule of classes. Please make yourself aware of these dates.

Emergency Notification and Procedures

Eagle Alert is UNT's official, campus-wide emergency notification system for emergency events, inclement winter weather closures, or Tornado Warnings. Eagle Alert allows UNT administrators to quickly contact campus community members by phone, text, and email. Eagle Alert will also post to the Eagle Alert Twitter. In addition to receiving direct messages by phone and email, the system includes a feature called desktop override that takes control over most UNT-owned computers in offices, classrooms, and public spaces on campus. During the test or in the event of an emergency, computer screens, presentation screens, and digital signs will display a full-screen alert.

UNT faculty, staff, and students are automatically enrolled in Eagle Alert. Individuals should remember, however, that if their personal contact information changes, they should go to my.unt.edu (students/faculty) or my.untsystem.edu (staff) to update their information. Instructions for updating your information can be found here.

Retention of Student Records

All university records must be retained and disposed of in accordance with state law and the university’s record retention schedule as approved by the Texas State Library and Archives Commission. Please read the full text of the UNT policy here.

Student Success

UNT endeavors to offer you a high-quality education and to provide a supportive environment to help you learn and grow. And, as a faculty member, I am committed to helping you be successful as a student. Here’s how to succeed at UNT: Show up. Find Support. Get advised. Be prepared. Get involved. Stay focused. To learn more about campus resources and information on how you can achieve success, go http://success.unt.edu/.

Incomplete Grade (I)

The grade of I (incomplete) is not given except for rare and very unusual emergencies, as per University guidelines. An I grade cannot be used to substitute for your poor performance in class. If you think you will not be able to complete the class satisfactorily, please drop the course.

Campus Closures

Should UNT close campus, it is your responsibility to check your official UNT e-mail account (EagleConnect) and Canvas to determine if your instructor plans to modify class activities, and how. This may include changing assignment due dates, rescheduling exams, holding online classes, etc.

Student Evaluation of Instruction

Student feedback is important and an essential part of participation in this course. The student evaluation of instruction is a requirement for all organized classes at UNT. The survey will be made available during weeks 13, 14 and 15 of the long semesters to provide students with an opportunity to evaluate how this course is taught. Students will receive an email from UNT SPOT Course Evaluations with the survey link. Students should look for the email in their UNT email inbox. Simply click on the link and complete the survey. For additional information, please visit the SPOT website.